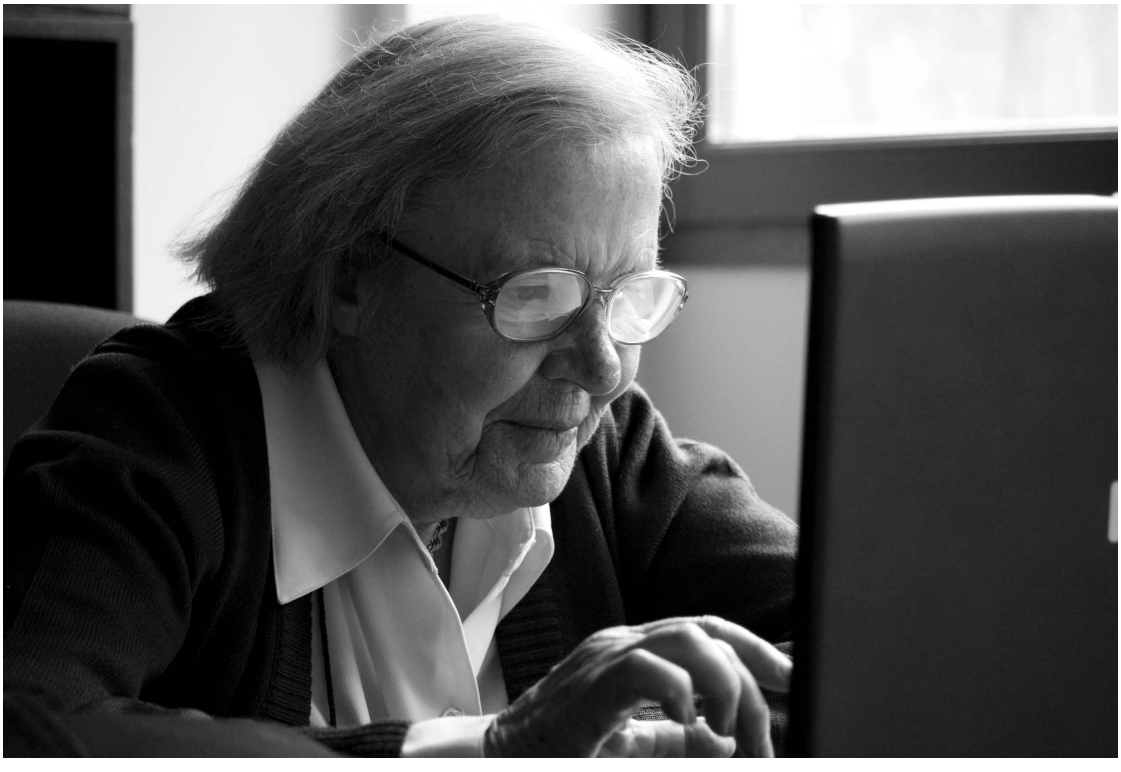
Comptes Rendus de l'Académie des sciences

1873-7234 (electronic)

Mécanique



Thematic issue / *Numéro thématique*

Recent advances in general relativity: an issue in memory of Yvonne Choquet-Bruhat / *Avancées récentes dans le domaine de la relativité générale : numéro en l'honneur d'Yvonne Choquet-Bruhat*

Guest Editor / *Rédacteur en chef invité*

Jérémie Szeftel

2025

Académie des sciences - Paris

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

CENTRE MERSENNE

Comptes Rendus
Mécanique

Objective of the journal

Comptes Rendus Mécanique is an international peer-reviewed electronic journal, covering all areas of mechanics and engineering science.
It publishes original research articles, review articles, historical perspectives, pedagogical texts or conference proceedings, of unlimited length, in English or in French and in as flexible a format as necessary (figures, associated data, etc.).
Comptes Rendus Mécanique has been published since 2020 with the centre Mersenne pour l'édition scientifique ouverte (Mersenne Center for open scientific publishing), according to a virtuous Diamond Open Access policy, free for authors (no author processing charges nor publishing fees) as well as for readers (immediate and permanent open access).

Editorial director: Étienne Ghys.

Editors-in-chief: Jean-Baptiste Leblond, Samuel Forest.

Associate editors: Olga Budenkova, Francisco Chinesta, Félix Darve, Florian Gosselin, Éric Lemarchand, Nicolas Moës, Léo Morin, Benoît Perthame, Guillaume Ribert, Géry de Saxcé, Emmanuel Villermaux.

Editorial secretary: Adenise Lopes.

About the journal

All journal's information, including the text of published articles, which is fully open access, is available from the journal website at https://comptes-rendus.academie-sciences.fr/mecanique/.

Author enquiries

For enquiries relating to the submission of articles, please visit this journal's homepage at https://comptes-rendus.academie-sciences.fr/mecanique/.

Contact

Académie des sciences
23, quai de Conti, 75006 Paris, France
Tel: (+33) (0)1 44 41 43 72
cr-mecanique@academie-sciences.fr

To cite this special issue: Jérémie Szeftel (ed). Recent advances in general relativity: an issue in memory of Yvonne Choquet-Bruhat. Comptes Rendus. Mécanique (2025), doi : 10.5802/crmeca.sp.1.

The articles in this journal are published under the license
Creative Commons Attribution 4.0 International (CC-BY 4.0)
https://creativecommons.org/licenses/by/4.0/deed.en

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 477-478
https://doi.org/10.5802/crmeca.289

Foreword / *Avant-propos*

Foreword to the special issue "Recent advances in general relativity: an issue in memory of Yvonne Choquet-Bruhat"

Avant-propos du numéro spécial « Avancées récentes dans le domaine de la relativité générale : numéro en l'honneur d'Yvonne Choquet-Bruhat »

Jérémie Szeftel [a]

[a] CNRS & Laboratoire Jacques-Louis Lions, Sorbonne université, France
E-mail: jeremie.szeftel@sorbonne-universite.fr

Yvonne Choquet-Bruhat (Photo credit: Geneviève Choquet)

The present volume is dedicated to the memory of Yvonne Choquet-Bruhat who recently passed away at 101 years old. She was the first woman elected at the French Academy of Sciences and a remarkable mathematician. She pioneered the field of mathematical general relativity with her seminal 1952 paper [1] on the local solvability of the Einstein equations. Her subsequent influential works on the constraint equations and on the evolution problem have left a deep impact on both theoretical and numerical developments. Her textbook [2] is widely viewed as a classical reference on the subject and an invaluable source of fruitful research directions. For more on the work of Yvonne Choquet-Bruhat, we refer the reader to the text "On Yvonne"

written by Piotr Chruściel in her honor for this special issue. We also recommend her memoirs [3] for the witty account of a woman mathematician at a time where they were very few.

The volume consists of a collection of reviews on various aspects of general relativity from a mathematical perspective, many of them originating from the work of Yvonne Choquet-Bruhat. The following topics are covered:

(1) "A survey on the positive mass theorem for asymptotically flat initial data sets" by Lan-Hsuan Huang
(2) "Gluing for the Einstein constraint equations" by Justin Corvino
(3) "The stability of Minkowski space and its influence on the mathematical analysis of general relativity" by Jacques Smulevici
(4) "The black hole stability problem" by Sergiu Klainerman
(5) "On the regularity problems of the Einstein equations" by Qian Wang
(6) "Weak Cosmic Censorship, trapped surfaces, and naked singularities for the Einstein vacuum equations" by Yakov Shlapentokh-Rothman
(7) "The Strong Cosmic Censorship conjecture" by Maxime Van de Moortel
(8) "Cosmology, the Big Bang and the BKL conjecture" by Hans Ringström
(9) "Burnett's conjecture in general relativity" by Arthur Touati
(10) "Some aspects of spectral and microlocal methods in general relativity" by Dietrich Häfner.

We hope that these surveys will allow the reader to gain insights into the latest developments of the field, as well as to appreciate the tremendous achievements that have been obtained along the way, starting with the pioneering contributions of Yvonne Choquet-Bruhat.

References

[1] Y. Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), pp. 141–225.
[2] Y. Choquet-Bruhat, *General Relativity and the Einstein Equations*, Oxford Mathematical Monographs, Oxford University Press: Oxford, New York, 2009.
[3] Y. Choquet-Bruhat, *Une mathématicienne dans cet étrange univers*, Odile Jacob: Paris, 2016.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 411-414
https://doi.org/10.5802/crmeca.272

Historical Commentary / *Commentaire historique*

On Yvonne

À propos d'Yvonne

Piotr T. Chruściel[a]

[a] University of Vienna, Faculty of Physics, Boltzmanngasse 5, A1090 Vienna, Austria
E-mail: piotr.chrusciel@univie.ac.at

Abstract. It is a pleasure and honor to present here a personal perspective on Yvonne Choquet-Bruhat's scientific work.

Résumé. C'est un plaisir et un honneur de présenter ici une perspective personnelle sur le travail scientifique d'Yvonne Choquet-Bruhat.

Keywords. Mathematical general relativity, Einstein equations, General relativistic Cauchy problems.

Mots-clés. Relativité générale mathématique, Équations d'Einstein, Problèmes de Cauchy en relativité générale.

Manuscript received 9 October 2024, accepted 11 October 2024.

To make this note less subjective, let me start by mentioning the citation list of Yvonne Choquet-Bruhat on MathSciNet, which is opened by her 2009 book

"General relativity and the Einstein equations"

published by Oxford University Press. (I had the pleasure to accompany Yvonne to the OUP offices when she was preparing her book for publication ☺.) In this list the book is followed by the milestone papers

"Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", Acta Math. 88 (1952), 141–225,

and

"Global aspects of the Cauchy problem in general relativity" with Robert Geroch, Commun. Math. Phys. 14 (1969), 329–335.

These last two papers are to mathematical relativity as what Banach's "Théorie des opérations linéaires" is to functional analysis.

Indeed, her brilliant proof, that Einstein equations possess a well posed Cauchy problem, is the foundation on which mathematical general relativity is built. Nowadays we take it for granted that our universe is described by evolution equations, with initial data prescribed on an initial data surface. The fact that this is the case for general relativity was not known until Yvonne's "Théorême d'existence..." [1]. There are several difficulties here. The first is the diffeomorphism invariance of the Einstein equations, which implies that solutions of Einstein equations can only be unique up to diffeomorphisms. The first discovery in [1] was, that this can be cured by introducing harmonic coordinates. The next difficulty is to prove that the resulting "harmonically-reduced" equations admit local solutions. Her paper provides the first proof of this for a large class of equations with the relevant structure. Last but not least, one needs to

show that a solution of the reduced equations provide a solution of the equations of real interest, namely the Einstein equations. It is shown in [1] how the general relativistic constraint equations can be used to solve this problem. A self-contained presentation of these constructions using modern techniques can be found in [2].

The local constructions of [1] immediately raise the question of existence and properties of global solutions of Einstein equations. The second milestone on which mathematical general relativity rests is the already-mentioned paper with Robert Geroch [3]. The paper unravels the fundamental role of the notion of global hyperbolicity in this context, and presents the proof that every set of, say vacuum, general relativistic initial data evolves to a unique spacetime which is maximal in the set of globally hyperbolic spacetimes inducing the given initial data. A nice conceptual simplification of the proof can be found in [4].

It turns out that this is not the end of the story: there exist well behaved initial data sets, such as initial data for the Kerr black holes or for the Bianchi IX cosmological models, for which the maximal globally hyperbolic developments can be extended beyond a Cauchy horizon in a non-unique way. The contents of the *cosmic censorship conjectures* is that this happens only in very special circumstances; more on this can be found in [5, 6]. To resolve this one needs to understand the global structure of generic solutions of the general relativistic Cauchy problem, which is not a minor task. Among the most significant achievements on this topic in the class of spatially compact spacetimes with one Killing vector are the results presented by Yvonne and Vince Moncrief in [7]. In this context it should be added that Yvonne Choquet-Bruhat's analysis of Fuchsian partial differential equations, reviewed in her already mentioned monograph [5], remains one of the best tools to explore the dynamics of cosmological spacetimes near singularities.

Needless to say, these are not the only contributions of Yvonne: MathSciNet lists 239 publications, including 11 books. One finds there seminal papers on the constraint equations in general relativity, including the development, with Demetrios Christodoulou, of the framework for analysing these equations on asymptotically flat manifolds [8]. These constraint equations constitute a set of nonlinear equations that has to be satisfied by the initial data, and one of the key goals of mathematical general relativity is to construct exhaustive families of their solutions, and to understand the properties of these. For many years a standard tool for this has been the "conformal method", some aspects of which have been intensively studied in the mathematics literature in the context of the Yamabe problem [9–11] mainly on compact manifolds. The paper [8] provides key tools to understand elliptic systems on asymptotically flat manifolds, including the general relativistic constraint equations.

One of the fundamental properties of asymptotically flat solutions of the constraint equations is positivity of mass [12, 13]. In a pioneering work with Marsden, Choquet-Bruhat proves positivity for near-Minkowskian initial data [14]. The arguments used there are used nowadays in state-of-the-art gluing theorems for the constraint equations [15, 16].

The general relativistic constraint equations on compact manifolds become quite a headache when matter fields transform in an inconvenient way under conformal transformations. An example is provided by the scalar field, which has been addressed by Yvonne and collaborators in [17]. More on the general relativistic constraint equations can be found in [18, 19], see also [20–24] and references therein.

Choquet-Bruhat's pioneering work with Daniel Bancel, "Existence, uniqueness, and local stability for the Einstein–Maxwell–Boltzman system", Comm. Math. Phys. 33 (1973), 83–96, provides the first rigorous treatment of the Boltzmann equation within Einsteinian gravity. Her prophetic "Ondes asymptotiques et approchées pour des systèmes d'équations aux dérivées partielles non linéaires" J. Math. Pures Appl. (9) 48 (1969), 117–158, has been given new life in recent papers [25, 26], which come close to settling Burnett's conjecture [27] concerning a class of high-frequency limits of solutions of Einstein's equations.

I had the pleasure and honor to coauthor a series of papers with Yvonne on the characteristic Cauchy problem in general relativity. There we present a variation of an approach of Rendall, perhaps making clearer the constraints character of some of Einstein equations. In [28] we observe that all the constraints can be reduced to linear equations, the only obstruction to global existence being positivity of a conformal factor. This should be compared to the much more intricate structure of the already-mentioned spacelike constraints. Natural extensions of this work can be found in [29, 30]. It should be mentioned that the role of the characteristic Cauchy problem in general relativity has been steadily growing, see [31–34].

Amongst the various honors received, Yvonne was the first female member of the French Academy of Sciences. She has been included in the list of the 85 original and important female contributors to the physics of the 20th Century on the web site URL https://web.archive.org/web/20161007090528/http://cwp.library.ucla.edu/exp.html. Together with James York, from Cornell University, she was awarded the 2003 Dannie Heineman Prize for Mathematical Physics,"for their separate as well as joint work in proving the existence and uniqueness of solutions to Einstein's gravitational field equations so as to improve numerical solution procedures with relevance to realistic physical solutions". The prize is presented in recognition of outstanding publications in the field of mathematical physics. In 2006 Yvonne Choquet-Bruhat delivered the Emmy Noether Lecture at the International Congress of Mathematicians in Madrid on the topic of "Mathematical problems in General Relativity". Her biography has been included on the web site 'Biographies of women mathematicians", URL https://mathwomen.agnesscott.org/women/bruhat.htm, next to Hypathia, Sofia Kovalevskaya, Emmy Noether, and Miryam Mirzakhani. A volume in her honor can be found on URL https://celebratio.org/ChoquetBruhat_Y/article/1108/.

References

[1] Y. Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), p. 141-225.

[2] H. Ringström, *The Cauchy problem in General Relativity*, ESI Lectures in Mathematics and Physics, European Mathematical Society (EMS), Zürich, 2009, MR 2527641 (2010j:83001).

[3] Y. Choquet-Bruhat, R. Geroch, "Global aspects of the Cauchy problem in general relativity", *Commun. Math. Phys.* **14** (1969), p. 329-335, MR MR0250640 (40 #3872).

[4] J. Sbierski, "On the existence of a maximal Cauchy development for the Einstein equations: a dezornification", *Ann. Henri Poincaré* **17** (2016), p. 301-329, MR 3447847.

[5] Y. Choquet-Bruhat, *General Relativity and the Einstein Equations*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2009, MR 2473363.

[6] P. T. Chruściel, *On Uniqueness in the Large of Solutions of Einstein Equations ("Strong Cosmic Censorship")*, Australian National University Press, Canberra, 1991.

[7] Y. Choquet-Bruhat, V. Moncrief, "Future global in time Einsteinian spacetimes with U(1) isometry group", *Ann. Henri Poincaré* **2** (2001), p. 1007-1064, MR 1877233.

[8] D. Christodoulou, Y. Choquet-Bruhat, "Elliptic systems in $H_{s,\delta}$ spaces on manifolds which are Euclidean at infinity", *Acta. Math.* **146** (1981), p. 129-150.

[9] S. Brendle, "Blow-up phenomena for the Yamabe equation", *J. Am. Math. Soc.* **21** (2008), p. 951-979, MR 2425176 (2009m:53084).

[10] M. A. Khuri, F. C. Marques, R. M. Schoen, "A compactness theorem for the Yamabe problem", *J. Differ. Geom.* **81** (2009), p. 143-196, MR 2477893 (2010e:53065).

[11] J. M. Lee, T. H. Parker, "The Yamabe problem", *Bull. Amer. Math. Soc. (N.S.)* **17** (1987), p. 37-91, MR MR888880 (88f:53001).

[12] E. Witten, "A simple proof of the positive energy theorem", *Commun. Math. Phys.* **80** (1981), p. 381-402.

[13] R. Schoen, S.-T. Yau, "The energy and the linear momentum of spacetimes in general relativity", *Commun. Math. Phys.* **81** (1981), p. 47-51.

[14] Y. Choquet-Bruhat, J. E. Marsden, "Solution of the local mass problem in general relativity", *Commun. Math. Phys.* **51** (1976), p. 283-296, MR 478215.

[15] Y. Mao, S.-J. Oh, Z. Tao, "Initial data gluing in the asymptotically flat regime via solution operators with prescribed support properties", preprint, 2023, [math.AP], https://arxiv.org/abs/2308.13031.

[16] S. Czimek, I. Rodnianski, “Obstruction-free gluing for the Einstein equations”, preprint, 2022, [gr-qc], https://arxiv.org/abs/2210.09663.
[17] Y. Choquet-Bruhat, J. Isenberg, D. Pollack, “The constraint equations for the Einstein-scalar field system on compact manifolds”, *Class. Quantum Grav.* **24** (2007), p. 809-828, MR 2297268.
[18] Y. Choquet-Bruhat, J. York, “The cauchy problem”, in *General Relativity* (A. Held, ed.), Plenum Press, New York, 1980, MR 583716 (82k:58028), p. 99-172.
[19] A. Carlotto, “The general relativistic constraint equations”, *Living Rev. Relativ.* **24** (2021), article no. 2.
[20] P. Hintz, “Gluing small black holes along timelike geodesics I: Formal solution”, preprint, 2023, [gr-qc], https://arxiv.org/abs/2306.07409.
[21] J. Corvino, “Scalar curvature deformation and a gluing construction for the Einstein constraint equations”, *Commun. Math. Phys.* **214** (2000), p. 137-189, MR MR1794269 (2002b:53050).
[22] J. Corvino, R. Schoen, “On the asymptotics for the vacuum Einstein constraint equations”, *J. Differ. Geom.* **73** (2006), p. 185-217, MR MR2225517 (2007e:58044).
[23] P. T. Chruściel, E. Delay, “On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications”, *Mém. Soc. Math. France.* **94** (2003), p. 1-103, MR MR2031583 (2005f:83008).
[24] J. Anderson, J. Corvino, F. Pasqualotto, “Multi-localized time-symmetric initial data for the Einstein vacuum equations”, *J. Reine Angew. Math.* **2024** (2024), no. 808, p. 67-110.
[25] C. Huneau, J. Luk, “High-frequency solutions to the Einstein equations”, *Class. Quantum Grav.* **41** (2024), no. 14, article no. 143002, MR 4777534.
[26] A. Touati, “The reverse Burnett conjecture for null dusts”, preprint, 2024, [math.AP], https://arxiv.org/abs/2402.17530.
[27] G. A. Burnett, “The high frequency limit in general relativity”, *J. Math. Phys.* **30** (1989), p. 90-96.
[28] Y. Choquet-Bruhat, P. T. Chruściel, J. M. Martín-García, “The light cone theorem”, *Class. Quantum Grav.* **26** (2009), p. 135011, MR 2515694 (2010g:53131).
[29] P. T. Chruściel, T.-T. Paetz, “The many ways of the characteristic Cauchy problem”, *Class. Quantum Grav.* **29** (2012), article no. 145006, MR 2949552.
[30] P. T. Chruściel, W. Cong, “Gluing variations”, *Class. Quantum Grav.* **40** (2023), article no. 165009.
[31] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2008, MR MR2488976 (2009k:83010).
[32] C. Kehle, R. Unger, “Gravitational collapse to extremal black holes and the third law of black hole thermodynamics”, preprint, 2022, [gr-qc], https://arxiv.org/abs/2211.15742.
[33] S. Klainerman, I. Rodnianski, “On the formation of trapped surfaces”, *Acta Math.* **208** (2012), p. 211-333, MR 2931382.
[34] D. Christodoulou, “The instability of naked singularities in the gravitational collapse of a scalar field”, *Ann. Math. (2)* **149** (1999), p. 183-217, MR MR1680551 (2000a:83086).

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 177-194
https://doi.org/10.5802/crmeca.270

Review article / *Article de synthèse*

A survey on the positive mass theorem for asymptotically flat initial data sets

Enquête sur le théorème de la masse positive pour les ensembles de données initiales asymptotiquement plats

Lan-Hsuan Huang [a]

[a] Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA
E-mail: lan-hsuan.huang@uconn.edu

Abstract. In honor of Yvonne Choquet-Bruhat's 100th birthday, we present this survey on the positive mass theorem. Originating from a conjecture in general relativity regarding the ADM mass, the positive mass theorem has significantly influenced geometry and analysis over the past four decades and continued to inspire new connections. We review seminal contributions as well as recent advances, and then we focus our discussions on the equality case and the counter-examples arising from pp-wave spacetimes.

Résumé. En l'honneur du 100e anniversaire d'Yvonne Choquet-Bruhat, nous présentons cette étude sur le théorème de la masse positive. Issu d'une conjecture en relativité générale concernant la masse de l'ADM, le théorème de la masse positive a influencé de manière significative la géométrie et l'analyse au cours des quatre dernières décennies et continue d'inspirer de nouvelles connexions. Nous passons en revue les contributions fondamentales ainsi que les avancées récentes, puis nous concentrons nos discussions sur le cas de l'égalité et les contre-exemples provenant des espaces-temps d'ondes pp.

Keywords. ADM mass, ADM energy-momentum, Asymptotically flat initial data set, Dominant energy condition, pp-wave.

Mots-clés. Masse ADM, Énergie-momentum ADM, Ensemble de données initiales asymptotiquement plates, Condition d'énergie dominante, Onde pp.

Funding. NSF DMS-2304966, NSF DMS-1928930.

Manuscript received 7 October 2024, accepted 10 October 2024.

1. Introduction

I am honored to contribute this survey on the positive mass theorem in celebration of Yvonne Choquet-Bruhat's 100th birthday. I had the chance to meet her in 2008 at the Mittag-Leffler Institute while I was a PhD student participating in the *Geometry, Analysis, and General Relativity* program. She visited briefly to meet her collaborators, including Piotr Chruściel and Jim Isenberg who were key participants of the program. During tea time, we gathered at a large oval table in the sunroom. As I was just starting my research career, her warmth and friendliness left a lasting impression. We are truly fortunate to have a pioneer like Choquet-Bruhat in the general relativity community. Her groundbreaking work and kindness are an amazing example for us all and continue to inspire us today.

Choquet-Bruhat made a fundamental contribution to the Cauchy problem, ensuring that the initial value problem in general relativity is well-posed. For more details, see her survey [1]. Her research laid the groundwork for the $3+1$ formulation of general relativity (ADM formalism), which is crucial for understanding how spacetime evolves and is used in both theoretical and numerical relativity. The ADM approach views spacetime as a series of evolving three-dimensional slices from an initial data set. It simplifies Einstein's field equations into constraint and evolution equations. It also introduces ADM energy and linear momentum, which measure the total mass of an asymptotically flat spacetime. These quantities, defined on the initial data set, are essential for understanding isolated gravitational systems and allow working with just the initial data and constraint equations. We refer the reader to her beautiful book [2].

The positive mass conjecture asserts that the ADM mass is non-negative (or, more precisely, that the ADM energy-momentum is future causal) under suitable energy conditions. Choquet-Bruhat and Marsden proved a very interesting case of the positive mass theorem, for metrics close to the Euclidean metric [3], building on an idea outlined by Brill and Deser [4]. The problem gained wider attention through questions posed by Geroch in his 1975 plenary address at the Joint Mathematical Meeting [5]. This conjecture became one of the early topics in mathematical general relativity and remains central to developments in geometric analysis and general relativity. The search for and affirmation of positivity conditions under various settings has had a significant impact on mathematics, influencing areas like the Yamabe problem in conformal geometry [6].

While there were other partial results toward the positive mass conjecture, the first groundbreaking result was obtained by Schoen and Yau in 1979 [7]. For an earlier history of this problem, we refer to the introduction of [7]. Not only did they completely settle the so-called Riemannian case, but the minimal surface technique they introduced also revolutionized the field. Since then, more general versions, higher-dimensional cases, and new alternative proofs have emerged, making the topic even more dynamic in recent years. Researchers have continued to develop innovative techniques and approaches, expanding the scope and relevance of the positive mass theorem. While the physically interesting case is three spatial dimensions, there are various reasons and motivations for seeking a general statement in higher dimensions. We will provide a more comprehensive list of progress below. For now, we present the statement.

Theorem 1 (Spacetime positive mass theorem). *Let $n \geq 3$ and (M, g, k) be an n-dimensional asymptotically flat initial data set that satisfies the dominant energy condition. Suppose either $3 \leq n \leq 7$ or M is spin. Then*

$$E \geq |P|,$$

where (E, P) is the ADM energy-momentum vector of (M, g, k).

The above theorem says that the ADM energy-momentum vector (E, P) is future-directed and causal. Thus, one can define its Lorentzian norm as the *ADM mass* $m = \sqrt{E^2 - |P|^2}$, which is a spacetime invariant.

The special case of Theorem 1 where $k \equiv 0$ is known as the Riemannian case (or the time-symmetric case). In this case, $P = 0$, and the dominant energy condition implies nonnegative scalar curvature. It is also conventional to denote the ADM energy E by the ADM mass m.

Theorem 2 (Riemannian positive mass theorem). *Let $n \geq 3$ and (M, g) be an n-dimensional asymptotically flat manifold with nonnegative scalar curvature. Suppose either $3 \leq n \leq 7$ or M is spin. Then $m \geq 0$. Furthermore, $m = 0$ if and only if (M, g) is isometric to Euclidean space.*

Schoen and Yau proved this case in dimension three in 1979 and 1981 [7, 8] by introducing minimal surface techniques and extended it to dimensions less than eight [9, 10]. They argued that if the ADM mass $m < 0$, then there exists a codimension 1 area-minimizing minimal

surface with the induced metric being asymptotically flat with zero mass. The area-minimizing property allows for a conformal factor that further transforms the induced metric to one with zero scalar curvature and negative mass, contradicting the Riemannian positive mass theorem in one dimension lower and ultimately conflicting with the Gauss–Bonnet Theorem when $n = 3$. Since the above argument does not address the case $m = 0$, Schoen and Yau gave a separate argument to show that if $m = 0$, then the manifold is Ricci flat and thus isometric to Euclidean space.

Lohkamp [11] observed a cut-off technique to simplify the asymptotics of an asymptotically flat manifold with negative ADM mass, showing that it can be compactified to a torus while maintaining positive scalar curvature. Consequently, the Riemannian positive mass theorem is equivalent to the torus rigidity theorem of Schoen and Yau [12]. In more recent years, there have been rapid advances and alternative proofs of the Riemannian positive mass theorem for three dimensions. Huisken and Ilmanen [13] used the inverse mean curvature flow, Li [14] employed the Ricci flow, Bray, Kazaras, Khuri, and Stern [15] considered level sets of harmonic functions and Bochner identity, and Agostiniani, Mazzieri, and Oronzio [16] used Green's function.

For the general case $k \not\equiv 0$ in dimension three, Schoen and Yau proved $E \geq 0$ by introducing Jang's equation [17] and reduce the case to the Riemannian positive mass theorem. Eichmair generalized the Jang equation argument and proved the $E \geq 0$ theorem in dimensions less than eight [18]. These results also show that if $E = 0$, then (M, g, k) can be isometrically embedded in Minkowski spacetime with the second fundamental form k.

For the result $E \geq |P|$ as stated in Theorem 1, Witten provided a novel proof in dimension three [19, 20], which directly generalizes to higher dimensions for spin manifolds [21, 22]. Eichmair, Huang, Lee, and Schoen [23] extended the minimal surface approach in the Riemannian case developed by Schoen and Yau by introducing marginally outer trapped surfaces (MOTS) to prove Theorem 1. They addressed the challenges posed by MOTS, which, unlike minimal surfaces, do not arise from a variation principle of the initial data set. In the same paper, they also provided an alternative approach, showing that the $E \geq 0$ theorem implies $E \geq |P|$ by a new density theorem. Recently, Hirsch, Kazaras, and Khuri [24] provided an alternative proof for the 3-dimensional case using level sets of spacetime harmonic functions. We will briefly review those approaches in Section 3.

Challenges in higher dimensions $n \geq 8$ for the positive mass theorem arise because minimal hypersurfaces and MOTS can have singularities in these dimensions, causing the dimensional reduction argument to potentially break down. Lohkamp proposed a program to address these issues in a series of papers [25–28]. For the Riemannian case, Schoen and Yau [29] introduced a different approach using minimal slicings. Additionally, there are perturbation arguments to avoid singularities of minimal hypersurfaces for dimensions $n = 8$ by Smale [30] and $n = 9, 10$ by Chodosh, Mantoulidis, and Schulze [31]. Separately, while we shall only discuss $n \geq 3$ in this survey, there is also a Riemannian positive mass theorem for $n = 2$; see [32].

Neither the spinor proof in [19, 20] nor the MOTS proof in [23] includes a complete proof to characterize the equality case $E = |P|$ of Theorem 1. The natural conjecture states that if $E = |P|$, then $E = |P| = 0$, and thus by $E = 0$ rigidity, (M, g) can be isometrically embedded into Minkowski space with induced second fundamental form k. This would signify that the ADM mass $m = \sqrt{E^2 - |P|^2} = 0$, uniquely characterizes Minkowski spacetime.

The conjecture is shown to be true in dimensions $n = 3, 4$, but surprisingly, in dimensions $n \geq 5$ it has the subtlety depending on asymptotic flat decay rates, to be specified in Section 4 below. For $n \geq 5$, an initial data set with $E = |P|$ that does not exhibit the optimal asymptotic decay rate can have $E = |P| \neq 0$ and exists as a Cauchy hypersurface in a larger class of spacetimes known as pp-waves (short for *plane-fronted waves with parallel rays*). Minkowski spacetime is a special case of pp-waves. Those pp-wave examples also give counter-examples to the Bartnik stationary conjecture, see [33, Section 2].

We summarize the current state of the art below.

Theorem 3 (Equality in the spacetime positive mass theorem). *Let $n \geq 3$ and (M, g, k) be an n-dimensional asymptotically flat initial data set that satisfies the dominant energy condition and has $E = |P|$.*

- *If either $n = 3, 4$ or (M, g, k) satisfies the optimal asymptotic decay rates*[1]*, then $E = |P| = 0$ and (M, g) is isometrically embedded into Minkowski spacetime with the induced second fundamental form k.*
- *If $n \geq 5$ and M is spin, then (M, g, k) is isometrically embedded into a pp-wave spacetime with the induced second fundamental form k.*

We can break down the equality theorem into two separate statements: The first statement is to show that $E = |P|$ implies $E = |P| = 0$, which is the part that requires the optimal asymptotics decay rate. The second statement is to use the fact that $E = 0$ to find an embedding of the initial data set into Minkowski space.

Witten sketched an idea for proving $E = |P|$ rigidity of spin manifolds in his 1981 article [19]. Ashtekar-Horowitz [34] and Yip [35] gave arguments for $n = 3$ under extra spacetime assumptions. A complete, rigorous proof was given in the work of Beig and Chruściel for $n = 3$ by a spinor argument [36] and by Chruściel and Maerten [37] to higher-dimensional spin manifolds under the optimal asymptotic decay rates for dimensions $n \geq 5$. For general initial data sets without the spin assumption, Huang and Lee used the method of Lagrangian Multipliers introduced by Bartnik [38] to prove the first statement (under similar optimal asymptotic decay rates if $n \geq 5$). They also gave an alternative proof of $E = 0$ rigidity [39]. Unlike the other proofs, their proof is self-contained in the sense that it does not depend on how one proves that $E \geq |P|$ theorem or $E \geq 0$ theorem. For $n = 3$, Hirsch and Zhang proved the equality theorem for $n = 3$ from the level sets method [40].

On the other hand, Huang and Lee [33] discovered those pp-wave counterexamples in dimensions $n > 8$. Specifically, they found asymptotically flat spacelike slices in a large class of $(n+1)$-dimensional pp-wave spacetimes that are not isometric to Minkowski spacetime, and those spacelike slices have $E = |P| \neq 0$. Hirsch and Zhang further improved these counterexamples to dimensions $n \geq 5$ and showed that a spin asymptotically flat initial data set which satisfies the dominant energy condition and $E = |P|$ must be embeded into a pp-wave [41].

Given that multiple survey articles already exist on the Riemannian positive mass theorem from different perspectives (e.g., [10, 42, 43]), along with a graduate textbook by Lee [44] on those topics, this survey will primarily focus on the spacetime positive mass theorem, with particular emphasis on the equality case. The outline of the survey paper is as follows: In Section 2, we set the stage and include a variational aspect of the ADM energy-momentum.

In Section 3, we briefly review the different proofs of positivity $E \geq |P|$. In Section 4, we discuss the pp-waves and the equality case $E = |P|$.

2. Definitions, notations, and basic facts

2.1. *Initial data sets, the dominant energy condition, and asymptotic flatness*

Let $n \geq 3$. An *initial data set* is an n-dimensional manifold U equipped with a Riemannian metric g and a symmetric $(0,2)$-tensor k. Loosely speaking, an initial data set consists the information needed to determine the evolution of spacetime by viewing the triple (U, g, k) as a hypersurface

[1] We say that (M, g, k) satisfies the *optimal asymptotic decay rates* if the asymptotics of (g, k) defined by (2) below hold for some $q > 0$ and $0 < \alpha < 1$ satisfying $q + \alpha > n - 2$. This is an additional assumption beyond the usual decay rate condition $q > n/2$ when $n \geq 5$.

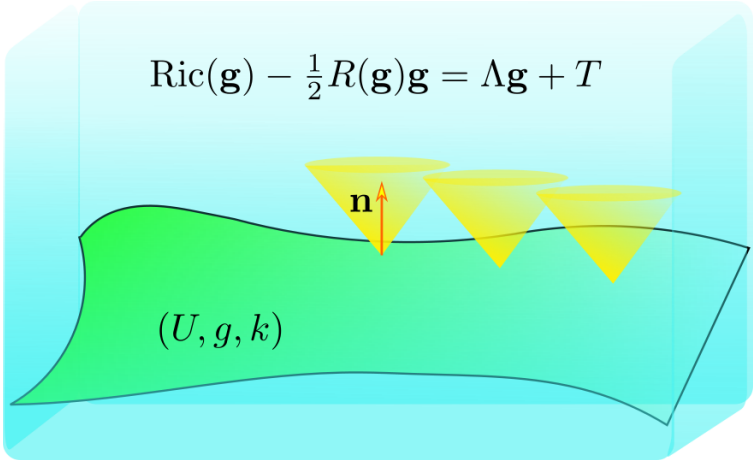


Figure 1. For an initial data set (U, g, k) embedded in a spacetime, the dominant energy condition on (U, g, k) states that $\mu \geq |J|_g$, where $\mu = T(\mathbf{n}, \mathbf{n})$ and $J_i = T(e_i, \mathbf{n})$, with $\mathbf{n}$ being the future-pointing unit normal to U and e_i tangent to U. This condition is equivalent to requiring that $T(\cdot, \mathbf{n})$ is future causal. Note that a spacetime satisfying the dominant energy condition asserts that $T(\mathbf{X}, \mathbf{Y}) \geq 0$ for all future causal vectors $\mathbf{X}$ and $\mathbf{Y}$, which is a stronger condition.

in spacetime $(\mathbf{N}, \mathbf{g})$ with the induced metric g and second fundamental form k. The *Einstein equation* of spacetime is given by

$$\mathrm{Ric}_{\mathbf{g}} - \tfrac{1}{2}R_{\mathbf{g}}\mathbf{g} = \Lambda\mathbf{g} + T,$$

where $\Lambda \in \mathbb{R}$ is the cosmological constant and T is the stress-energy tensor describing a matter field. In this article, we shall only discuss the case $\Lambda = 0$, but note that there are also many progresses on the positive mass theorem for the case $\Lambda < 0$; see, for example, [45–50]. A spacetime is called *vacuum* if T is identically zero. As an example of a non-vacuum matter field, the stress-energy tensor of a perfect fluid is given by $T = p\mathbf{g} + (\rho + p)\mathbf{v} \otimes \mathbf{v}$, where ρ is the energy density, p is the pressure, and $\mathbf{v}$ is the fluid velocity. We do not assume any specific matter fields here; the results apply to any matter field, as long as an energy condition is satisfied, unless otherwise specified.

On an initial data set, one can define the *mass density* μ and the *current density* J by

$$\begin{aligned} \mu &= \tfrac{1}{2}\left(R_g - |k|_g^2 + (\mathrm{tr}_g k)^2\right) \\ J &= \mathrm{div}_g k - d(\mathrm{tr}_g k). \end{aligned}$$

The first equation is referred to as the *Hamiltonian constraint* and the second is the *momentum constraint*. We denote the *constraint map* by $\Phi(g, k) := (\mu, J)$.

We say that (U, g, k) satisfies the *dominant energy condition* if $\mu \geq |J|_g$. If (U, g, k) is *embedded in a spacetime* $(\mathbf{N}, \mathbf{g})$, meaning that (U, g) is isometrically embedded in to $\mathbf{N}$ with the induced metric k, then the dominant energy condition is equivalent to requiring $T(\mathbf{X}, \mathbf{n}) \geq 0$ for all future causal vectors $\mathbf{X}$, where $\mathbf{n}$ is the future-directed normal to the hypersurface. The initial data set (U, g, k) is said to be *vacuum* if $\mu = 0, J = 0$. See Figure 1.

When considering the perturbation of the dominant energy condition, it turns out more effective to use the *modified constraint map* $\overline{\Phi}|_{(g,k)}$ *at* (g, k), introduced by Corvino and Huang [51]. The modified operator $\overline{\Phi}|_{(g,k)}$ is defined on initial data (γ, τ) as follows:

$$\overline{\Phi}_{(g,k)}(\gamma, \tau) = \Phi(\gamma, \tau) + \left(0, \tfrac{1}{2}\gamma \cdot J\right), \tag{1}$$

where J is the current density of g and $(\gamma\cdot J)^i := g^{ij}\gamma_{jk}J^k$. A fundamental property of the modified operator is that the dominant energy condition is preserved under perturbation. More precisely, assume (g,k) satisfies the dominant energy condition $\mu \geq |J|_g$ in M. Suppose (γ,τ) is an initial data set with $|\gamma - g|_g < 3$ in M and

$$\overline{\Phi}|_{(g,k)}(\gamma,\tau) = \overline{\Phi}|_{(g,k)}(g,k).$$

Then (γ,τ) also satisfies the dominant energy condition.

The seminal work of Choquet-Bruhat says that if (U,g,k) is a vacuum initial data set, then there exists a unique vacuum spacetime development $(\mathbf{N},\mathbf{g})$ that evolves from (U,g,k). However, note that in general an initial data set satisfying the dominant energy condition may not have a spacetime development satisfying an energy condition, or if it does, the spacetime development may not be unique.

For analysis on asymptotically flat manifolds, it is convenient to use the weighted Hölder spaces, defined below. Let B be a closed ball in $\mathbb{R}^n$ centered at the origin, and consider the exterior region $\mathbb{R}^n \setminus B$. For any $\ell = 0,1,2,\ldots$, $\alpha \in (0,1)$, $q \in \mathbb{R}$, we define the *weighted Hölder space* $C^{\ell,\alpha}_{-q}(\mathbb{R}^n \setminus B)$ as the space of functions f on $\mathbb{R}^n \setminus B$ such that

$$\|f\|_{C^{\ell,\alpha}_{-q}(\mathbb{R}^n\setminus B)} := \sum_{|\alpha|\leq \ell} \sup_x \left||x|^{|\alpha|+q}\partial^\alpha f(x)\right| + \sum_{|\alpha|=\ell} \left[|x|^{|\alpha|+q}\partial^\alpha f(x)\right]_\alpha < \infty.$$

Now suppose that M is a manifold such that there is a compact set $K \subset M$ and a diffeomorphism $M \setminus K \cong \mathbb{R}^n \setminus B$. Then one can define $C^{\ell,\alpha}_{-q}(M)$ by choosing an atlas for M that consists of the diffeomorphism $M \setminus K \cong \mathbb{R}^n \setminus B$ together with finitely many precompact charts, and then using the $C^{\ell,\alpha}_{-q}(\mathbb{R}^n \setminus B)$ norm on the noncompact chart while using the $C^{\ell,\alpha}$ norm on the other ones. The definition of $C^{\ell,\alpha}_{-q}(M)$ also extends to tensors simply by considering their components with respect to these coordinate charts.

We say that an initial data set (M,g,k) is *asymptotically flat with decay rate* q if M is a complete manifold and there is a compact set $K \subset M$ and a diffeomorphism $M \setminus K \cong \mathbb{R}^n \setminus B$ for some closed ball $B \subset \mathbb{R}^n$ such that

$$(g-\delta,k) \in C^{2,\alpha}_{-q}(M) \times C^{1,\alpha}_{-1-q}(M) \tag{2}$$

where δ is a smooth symmetric $(0,2)$-tensor that coincides with the Euclidean inner product on $M \setminus K \cong \mathbb{R}^n \setminus B$. In addition, we assume $\mu, J \in L^1(M)$, and sometimes it is convenient to make a stronger assumption that $\mu, J \in C^{0,\alpha}_{-n-q_0}(M)$ for some $q_0 > 0$. Throughout the article, we assume $q > (n-2)/2$.

2.2. *ADM energy and linear momentum, Regge–Teitelboim Hamiltonian, and Killing initial data*

Assume (M,g,k) is asymptotically flat with decay rate $q > (n-2)/2$ in the above sense. We define the *ADM energy* E and the *ADM momentum* P as

$$\begin{aligned} E(g,k) &= \frac{1}{2(n-1)\omega_{n-1}} \lim_{r\to\infty} \int_{|x|=r} \sum_{i,j=1}^n \left(\frac{\partial g_{ij}}{\partial x^i} - \frac{\partial g_{ii}}{\partial x^j}\right) \nu_0^j \, \mathrm{d}\sigma_0 \\ P_i(g,k) &= \frac{1}{(n-1)\omega_{n-1}} \lim_{r\to\infty} \int_{|x|=r} \sum_{j=1}^n (k_{ij} - (\mathrm{tr}_g k) g_{ij}) \nu_0^j \, \mathrm{d}\sigma_0 \qquad i = 1,2,\ldots,n. \end{aligned}$$

Here, the integrals are computed in the coordinate chart $M \setminus K \cong_x \mathbb{R}^n \setminus B$, $\nu_0^j = x^j/|x|$, σ_0 is the $(n-1)$-dimensional Euclidean Hausdorff measure, and ω_{n-1} is the volume of the standard unit sphere in $\mathbb{R}^n$.

Notice that if the decay rate $q > n-2$, then the integrand in the above expressions decays too quickly, so $E = 0$ and $P = 0$ trivially. On the other hand, a decay rate $q > (n-2)/2$ is sufficient

to ensure that the ADM energy-momentum is well-defined in the following sense. While the ADM energy and linear momentum are defined using a specific asymptotically flat coordinate chart, they can be shown to be coordinate invariant, depending only on the structure at infinity, see [21]. Furthermore, although the ADM energy and linear momentum are defined entirely on the initial data set, the Lorentzian norm $-E^2+|P|^2$ is a spacetime invariant. It means that, while the ADM energy-momentum vectors of two asymptotically flat initial data sets embedded in the same spacetime may differ, they have the same Lorentzian norm, see [52].

To understand heuristically how the decay rate $q > (n-2)/2$ plays a role, we expand the constraint equations in asymptotically flat coordinates:

$$\mu = \frac{1}{2}(R_g - |k|^2 + (\mathrm{tr}_g k)^2) = \frac{1}{2}\sum_{i,j}\frac{\partial}{\partial x^j}\left(\frac{\partial g_{ij}}{\partial x^i} - \frac{\partial g_{ii}}{\partial x^j}\right) + O(|x|^{-2-2q})$$
$$J_i = (\mathrm{div}_g k)_j - (\mathrm{tr}_g k)_{,j} = \sum_i \frac{\partial}{\partial x^i}(k_{ij} - (\mathrm{tr}_g k) g_{ij}) + O(|x|^{-2-2q}).$$

Observe that the ADM energy-momentum arises from the flux integrals of the divergence terms on the right-hand sides. The decay rate $q > (n-2)/2$ ensures that the exponent of the error terms, $-2-2q < -n$, decays rapidly enough to be integrable on the asymptotically flat end. Together with the assumption that both μ and J are integrable, the flux integrals over the coordinate sphere have a limit as the radius of the sphere going to infinity.

The following density theorem states that initial data sets with the decay rate $q = n-2$ are dense among those with the weaker decay rates $q > (n-2)/2$, and that the ADM energy-momentum varies continuously. This has several applications. For example, to prove the positivity of the ADM mass as stated in Theorem 1, it suffices to verify the inequality for an asymptotically flat initial data set with a decay rate $q = n-2$. For more details about the density theorem, see, for example, [23, Section 6].

Theorem 4 (Density theorem (loosely stated)). *Let (M, g, k) be an n-dimensional asymptotically flat initial data set at the decay rate $q > (n-2)/2$ satisfying the dominant energy condition. For any $\epsilon > 0$, there exists an asymptotically flat initial data set $(M, \bar{g}, \bar{k})$ at the decay rate $q = n-2$ such that $(\bar{g}, \bar{k})$ satisfies the dominant energy condition, is ϵ-close to (g, k) in $C^{2,\alpha}_{-q}(M) \times C^{1,\alpha}_{-1-q}(M)$, and $|(\bar{E}, \bar{P}) - (E, P)| < \epsilon$.*

Furthermore, we can arrange $(M, \bar{g}, \bar{k})$ to have either the strict dominant energy condition $\bar{\mu} > |\bar{J}|_{\bar{g}}$ everywhere in M or to be vacuum $\bar{\mu} = |\bar{J}|_{\bar{g}} = 0$ outside a large compact subset of M.

The ADM energy-momentum has a variational characterization.

Definition 5 (Regge–Teitelboim Hamiltonian). *Let M be an n-dimensional manifold that can carry an asymptotically flat initial data set. Let (f_0, X_0) be a pair of a function and a vector field on M (which we will often call a* lapse-shift pair*) such that (f_0, X_0) is smooth and is equal to a constant $(a, b) \in \mathbb{R} \times \mathbb{R}^n$ in the exterior coordinate chart for $M \setminus K$. We define the* Regge–Teitelboim Hamiltonian *$\mathscr{H}_{\mathrm{RT}}$ corresponding to (f_0, X_0) by, for asymptotically flat initial data set (g, k)*

$$\mathscr{H}_{\mathrm{RT}}(g, k) = (n-1)\omega_{n-1}\left[aE(g, k) + b \cdot P(g, k)\right] - \int_M \langle \Phi(g, k), (f_0, X_0)\rangle_g \, \mathrm{d}\mu_g. \tag{3}$$

In the case of vacuum initial data sets, as originally considered by Regge and Teitelboim [53], the Hamiltonian can recover the energy E and the linear momentum by choosing the lapse-shift pair to be asymptotic to the translation vector fields of Minkowski spacetime. Specifically, $(f, X) \to (1, 0)$ yields the energy E, while $(f, X) \to (0, \partial/\partial x^i)$ yields P_i, the i-th component of the linear momentum. The task of "minimizing" the ADM energy-momentum (E, P) is transformed

into minimizing the functional (3) as follows. Suppose the positivity of mass holds for vacuum (g,k), i.e., $E(g,k) \geq |P(g,k)|$. Then, for any future causal vector (a,b), we have

$$aE(g,k) + b \cdot P(g,k) \geq aE(g,k) - |b||P(g,k)| \geq (a - |b|)|P(g,k)| \geq 0 \tag{4}$$

with equality if and only if $E(g,k) = |P(g,k)|$ and (a,b) is a constant multiple of $(E(g,k), P(g,k))$. This is just a general fact that for non-zero future causal vectors $\mathbf{X}$ and $\mathbf{Y}$, their Lorentzian inner product $-\langle \mathbf{X}, \mathbf{Y} \rangle \geq 0$ with equality if and only if $\mathbf{X}, \mathbf{Y}$ are both null and $\mathbf{X}$ is a constant multiple of $\mathbf{Y}$.

With the minimization task in mind, we compute the first variation of $\mathscr{H}_{\mathrm{RT}}$: for vacuum (g,k), we have

$$D\mathscr{H}\big|_{(g,k)}(h,w) = -\int_M (h,w) \cdot D\Phi|^*_{(g,k)}(f_0, X_0)\, \mathrm{d}\mu_g$$

for all compactly supported (h,w), where $D\Phi|^*_{(g,k)}$ denotes the L^2-formal adjoint equation of the linearization $D\Phi|_{(g,k)}$ at (g,k). The adjoint equation $D\Phi|^*_{(g,k)}$ has geometric significance in the vacuum case, as discovered by Moncrief [54].

Theorem 6. *Let (U,g,k) be a vacuum initial data set and suppose that there exists a nontrivial lapse-shift pair (f,X) on U solving*

$$D\Phi|^*_{(g,k)}(f,X) = 0.$$

Then the vacuum spacetime development of (U,g,k) admits a unique global Killing vector field $\mathbf{Y}$ such that $\mathbf{Y} = f\mathbf{n} + X$ along U, where $\mathbf{n}$ is the future unit normal to U.

Conversely, given a vacuum spacetime equipped with a global Killing vector field $\mathbf{Y}$ and a spacelike hypersurface U with induced initial data (g,k), if we decompose $\mathbf{Y} = f\mathbf{n} + X$ along U, then the lapse-shift pair (f,X) must lie in the kernel of $D\Phi|^*_{(g,k)}$.

In light of the above theorem, a solution (f,X) to $D\Phi|^*_{(g,k)}(f,X) = 0$ is referred to as the *Killing initial data.*

An analogous statement for a non-vacuum case was established by Huang and Lee [33, Theorem 6] using the modified constraint operator. A spacetime $(\mathbf{N}, \mathbf{g})$ is said to be a *null perfect fluid* with velocity $\mathbf{v}$ and pressure p if $\mathbf{v}$ is either future null or zero at each point and the Einstein tensor takes the form:

$$\mathrm{Ric}_{\mathbf{g}} - \tfrac{1}{2} R_{\mathbf{g}} \mathbf{g} = p\mathbf{g} + \mathbf{v} \otimes \mathbf{v}. \tag{5}$$

Theorem 7. *Let (U,g,k) be an initial data set satisfying the dominant energy condition. Assume there exists a nontrivial lapse-shift pair (f,X) on U solving the system*

$$D\overline{\Phi}|^*_{(g,k)}(f,X) = 0 \quad \textit{and} \quad fJ + |J|_g X = 0. \tag{6}$$

and assume that f is nonvanishing in U. Then the following holds:

(1) *The dominant energy scalar $\sigma(g,k) := \mu - |J|_g$ is constant on U.*
(2) *(U,g,k) is embedded inside a null perfect fluid spacetime $(\mathbf{N}, \mathbf{g})$ that satisfies the spacetime dominant energy condition and admits a global Killing vector field $\mathbf{Y}$ equal to $f\mathbf{n} + X$ along U, where $\mathbf{n}$ is the future unit normal to U.*

There is also a converse statement.

We remark that the system (6) is closely related to the improvability of the dominant energy condition. See [33].

We refer to a solution (f,X) to $D\overline{\Phi}|^*_{(g,k)}(f,X) = 0$ as the *modified* Killing initial data. These solutions play an important role in the equality theorem of the positive mass theorem; see Section 4. An overall strategy for proving the equality theorem is to show that the initial data set achieving equality has one or more modified Killing initial data.

3. The proofs of positivity $E \geq |P|$

We outline the currently available proofs of the positivity theorem, Theorem 1.

3.1. *Marginally outer trapped hypersurfaces*

We briefly revisit the minimal hypersurface proof of the Riemannian positive mass theorem, originally developed by Schoen and Yau. This proof uses induction on the dimension $3 \leq n \leq 7$ and proceeds by contradiction. Assume there exists an asymptotically flat Riemannian manifold (M, g) with nonnegative scalar curvature and negative mass $m < 0$. A density argument allows us to assume (M, g) has harmonic asymptotics and positive scalar curvature. These conditions imply that the coordinate planes $x^n = \pm\Lambda$ act as barriers for minimal hypersurfaces for sufficiently large Λ.

Consider an $(n-1)$-dimensional vertical cylinder ∂C_ρ with a large radius ρ. For each $h \in [-\Lambda, \Lambda]$, there exists an area-minimizing hypersurface $\Sigma_{\rho,h} \subset C_\rho$ whose boundary is the height h sphere on ∂C_ρ. If $n \leq 7$, this hypersurface is smooth. Each $\Sigma_{\rho,h}$ lies between the hyperplanes $x^n = \pm\Lambda$. The area of $\Sigma_{\rho,h}$ is minimized by some $h_\rho \in (-\Lambda, \Lambda)$. This "height-picking" step is crucial in dimensions $n \geq 4$ because the corresponding surface Σ_{ρ,h_ρ} satisfies

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_{t=0} \operatorname{area}(\Phi_t(\Sigma_{\rho,h_\rho})) \geq 0$$

for any variation Φ_t that is a vertical translation along ∂C_ρ. Without the height-picking step, the above inequality only holds for compactly supported variations Φ_t, which is not sufficient. (When $n = 3$, one can use the so-called logarithmic cut-off trick to approximate variations that are vertical translations along ∂C_ρ by compactly supported variations. However, this trick does not work in higher dimensions.) A smooth subsequential limit Σ_∞ of Σ_{ρ,h_ρ} as $\rho \to \infty$ is an $(n-1)$-dimensional asymptotically flat manifold with zero energy and is a stable minimal hypersurface of M.

For $n = 3$, the stability of Σ_∞ and its asymptotics are incompatible with the Gauss–Bonnet theorem. In higher dimensions, one can construct a conformal factor that changes the metric on Σ_∞ to one that is asymptotically flat with zero scalar curvature and negative mass, thus violating the Riemannian positive mass theorem in dimension $n-1$.

Our approach to the spacetime positive mass theorem generalizes the minimal hypersurface proof using marginally outer trapped hypersurfaces (MOTS). A hypersurface Σ is a MOTS if $\theta = 0$, where $\theta = H_\Sigma + \operatorname{tr}_\Sigma k$ is the expansion scalar. In the Riemannian case, θ reduces to the mean curvature.

Let (M, g, k) be an n-dimensional asymptotically flat initial data set satisfying the dominant energy condition $\mu \geq |J|$ and $E < |P|$. By our density theorem, Theorem 4, we may assume (M, g, k) has harmonic asymptotics and satisfies the strict dominant energy condition $\mu > |J|$. Furthermore, we assume P points in the vertical direction $-\partial_n$. The harmonic asymptotics and $E < |P|$ assumptions imply that the planes $x^n = \pm\Lambda$ act as barriers for MOTS for sufficiently large Λ. Consider an $(n-1)$-dimensional vertical cylinder ∂C_r. For each $h \in [-\Lambda, \Lambda]$, the existence theorem of MOTS [55] guarantees the existence of a MOTS $\Sigma_{r,h}$ with boundary equal to the height h sphere on ∂C_r. This MOTS is smooth if $n < 8$ and lies between the planes $x^n = \pm\Lambda$, being stable in the sense of MOTS with boundary. See Figure 2.

A major challenge, distinct from the minimal surface approach, is that MOTS do not arise from a variational principle. Thus, there is no obvious alternative to the height-picking step of

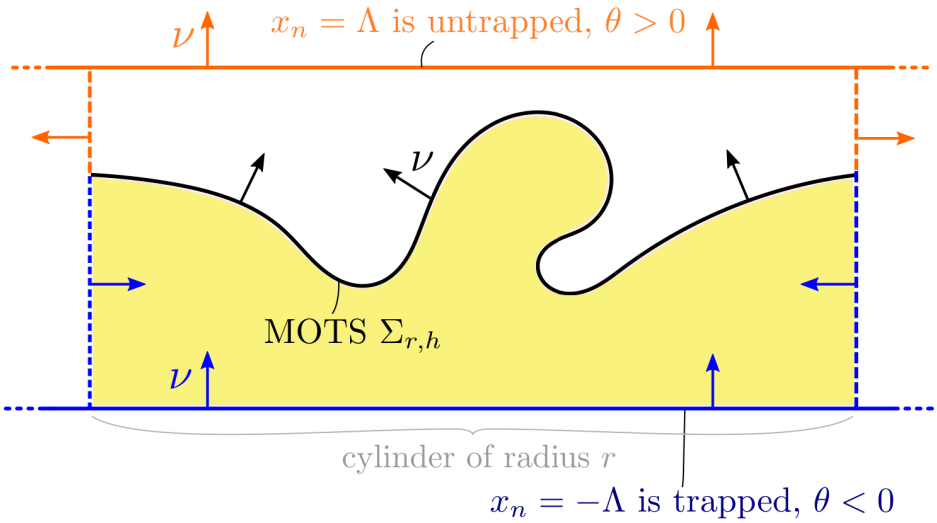


Figure 2. Under the (contradictory) assumption $E < |P|$, two coordinate hyperplanes $x^n = \Lambda$ and $x^n = -\Lambda$, together with the lateral side of a cylinder of large radius, provide barriers for the existence of a MOTS Σ_r with prescribed boundary.

minimizing the area with respect to the height h. To overcome this, we introduce a new functional $\mathscr{F}$ on hypersurfaces with boundary on ∂C_r such that for some $h_r \in (-\Lambda, \Lambda)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}h}\bigg|_{h=h_r} \mathscr{F}(\Sigma_{r,h}) \geq 0.$$

This inequality, along with the MOTS stability, parallels the area inequality in the Riemannian case. Extracting a smooth subsequential limit Σ_∞ of Σ_{r,h_r} as $r \to \infty$, we find that Σ_∞ is an $(n-1)$-dimensional asymptotically flat manifold with zero energy and is a stable MOTS in M. We can construct a conformal factor that changes the metric on Σ_∞ to one with zero scalar curvature and decreases the energy of Σ_∞, thus violating the Riemannian positive mass theorem in dimension $n-1$.

3.2. *Boost argument*

The positive mass theorem $E \geq |P|$ can also be derived from the positive energy theorem $E \geq 0$ using the following reduction argument. Assume that $0 < E < |P|$. Fix $\theta \in (0,1)$ such that $E' := (E - \theta|P|)/(1-\theta^2) < 0$. Using the density theorem, Theorem 4, we may assume that (M, g, k) satisfies $\mu = 0$ and $J = 0$ outside a large compact set. According to [56], there exists a vacuum spacetime development of the asymptotically flat end of (M, g, k) in which the end of (M, g, k) can be deformed to a boosted slice (of slope θ) that has energy E'. The deformed initial data (M', g', k') satisfies the conditions of the $E \geq 0$ theorem, leading to a contradiction. This reduction is well known in the mathematical relativity community when the initial data set is assumed to be a spacelike slice of an asymptotically flat spacetime; see, e.g., [57].

3.3. *Spinor proof*

Witten's proof is notable for its use of spinors, offering a different perspective on the positive mass theorem. A spinor field ψ on a manifold is a section of a spinor bundle, which can be roughly

thought of as a field of "spinor-valued" functions. For a 3-dimensional initial data set (M, g, k), M can be given a spin structure. In higher dimensions, this method requires M has a spin structure, which allows us to use spinor fields ψ on M. For convenience, we consider (M, g, k) embedded in a spacetime $(\mathbf{N}, \mathbf{g})$ and choose an orthonormal frame $\{e_0, e_1, \dots, e_n\}$ where e_0 is the timelike vector. One can also use a framework that is a purely initial data perspective. Define a new connection $\tilde{\nabla}_i = \nabla_i + (1/2)k_{ij}e_je_0$ and the Dirac operator $\not{D} = e_i \cdot \tilde{\nabla}_i = e_i \cdot \nabla_i + (1/2)k_{ij}e_ie_je_0$ where ∇_i is the induced connection on M. The Witten equation states

$$\not{D}^2\psi = \tilde{\nabla}^*\tilde{\nabla}\psi + \tfrac{1}{2}(\mu + Je_0)\cdot\psi.$$

For any ψ asymptotic to a constant spinor ψ_0 at infinity, integrating the above Witten's equation and analyzing the boundary term

$$\begin{aligned}\int_M |\tilde{\nabla}\psi|^2 - |\not{D}\psi|^2 + \tfrac{1}{2}\langle\psi, (\mu + Je_0)\cdot\psi\rangle \,\mathrm{d}\mu \\ = \lim_{r\to\infty}\int_{|x|=r}\langle\psi, \tilde{\nabla}_\nu\psi + \nu\cdot\not{D}\psi\rangle\,\mathrm{d}\sigma = \tfrac{1}{2}(n-1)\omega_{n-1}\left(|\psi_0|^2E + \langle\psi_0, P_je_je_0\cdot\psi_0\rangle\right)\end{aligned} \tag{7}$$

where ν is the outward unit normal. By choosing ψ to solve $\not{D}\psi = 0$ with ψ asymptotic to a certain constant spinor ψ_0 such that $\langle\psi_0, P_je_je_0\cdot\psi_0\rangle = -|\psi_0|^2|P|$, one gets

$$\tfrac{1}{2}(n-1)\omega_{n-1}|\psi_0|^2(E - |P|) = \int_M |\tilde{\nabla}\psi|^2 + \tfrac{1}{2}\langle\psi, (\mu + Je_0)\cdot\psi\rangle\,\mathrm{d}\mu \geq 0$$

where we use the dominant energy condition in the last inequality.

The argument also directly leads the $E = 0$ rigidity. Since $E = 0$ implies $|P| = 0$, for any constant spinor ψ_0 at infinity, one can let the harmonic spinor ψ asymptotically approach ψ_0 in (7) to obtain $\tilde{\nabla}\psi = 0$, meaning that ψ is, in fact, a parallel spinor. One then argue that the spacetime metric is flat along M. See [20] for more details.

3.4. *Level set method using spacetime harmonic functions*

We provide an overview of the proof by Hirsch, Kazaras, and Khuri [24]. Additionally, a recent survey by Bray, Hirsch, Kazaras, Khuri, and Zhang [58] discusses the level set method applied to positive mass theorems and other applications.

Let (Ω, g, k) be a 3-dimensional compact initial data set. Consider a scalar-valued function u. For simplicity, we assume that u has no critical points. Let $\Sigma_t = u^{-1}(t)$ be a level set of u, and denote the unit normal to Σ_t by $\eta = \nabla u/|\nabla u|$. By applying the Gauss equation to the level set Σ_t and using $A(e_\alpha, e_\beta) = |\nabla u|^{-1}\nabla^2 u(e_\alpha, e_\beta)$, we obtain

$$|\nabla u|\mathrm{Ric}(\eta,\eta) = \tfrac{1}{2}|\nabla u|(R_g - R_\Sigma) + |\nabla u|^{-1}\left(|\nabla|\nabla u||^2 - \tfrac{1}{2}|\nabla^2 u|^2 + \tfrac{1}{2}(\Delta u)^2 - (\Delta u)(\nabla^2 u)(\eta,\eta)\right). \tag{8}$$

We have the modified Bochner identity:

$$\Delta|\nabla u| = \frac{1}{|\nabla u|}\left(|\nabla^2 u|^2 + \langle\nabla u, \nabla\Delta u\rangle\right) + |\nabla u|\mathrm{Ric}(\eta,\eta) - |\nabla u|^{-1}|\nabla|\nabla u||^2.$$

Equating the previous two identities gives

$$\Delta|\nabla u| = \frac{1}{|\nabla u|}\langle\nabla u, \nabla\Delta u\rangle + \frac{1}{2}|\nabla u|\left[R_g - R_\Sigma + \frac{|\nabla^2 u|^2 + (\Delta u)^2 - 2|\nabla u|(\Delta u)\nabla^2 u(\eta,\eta)}{|\nabla u|^2}\right]. \tag{9}$$

So far, the identity holds for any function u. It was observed in [24] that this identity takes a more convenient form when u satisfies the *spacetime harmonic equation* $\Delta u + (\mathrm{tr}_g k)|\nabla u| = 0$, as motivated below. Suppose the initial data set (Ω, g, k) is embedded in a spacetime $(\mathbf{N}, \mathbf{g})$ with the unit normal $\mathbf{n}$. Then the Hessian operator ∇^2 of $\mathbf{g}$ restricted on Ω is given by

$$\nabla^2 u(e_i, e_i) = \nabla^2 u(e_i, e_j) + k_{ij}\mathbf{n}(u).$$

Since $\mathbf{n}(u)$ involves the spacetime vector $\mathbf{n}$ and is not determined on the initial data set (M, g, k), it turns out to be effective to require $\mathbf{n}(u) = |\nabla u|$; namely, the spacetime vector ∇u is null. By taking the trace, it gives the spacetime harmonic equation.

Substituting u, which solves the spacetime harmonic equation, into (9) yields

$$\Delta|\nabla u| = \frac{1}{2}|\nabla u|\left[2\mu - R_\Sigma + \frac{\left|\nabla^2 u + k|\nabla u|\right|^2}{|\nabla u|^2}\right] - \langle \nabla u, \nabla(\mathrm{tr}_g k)\rangle - \langle \nabla^2 u, k\rangle.$$

Consider the identity on an asymptotically flat initial data set (M, g, k) and let Ω be a large coordinate sphere. Consider u with the asymptotics $u(x) \to a \cdot x$ as $|x| \to \infty$ where $a \in \mathbb{R}^3$ and $|a| = 1$. Integrating the previous identity over Ω and applying the coarea formula gives the following integral formula:

$$\int_{\partial\Omega} (\nu(|\nabla u|) + k(\nabla u, \nu))\, \mathrm{d}\sigma = \frac{1}{2}\int_{\underline{u}}^{\overline{u}} \int_{\Sigma_t} \left(2\mu + 2J(\eta) - R_{\Sigma_t} + \frac{\left|\nabla^2 u + k|\nabla u|\right|^2}{|\nabla u|^2}\right) \mathrm{d}\sigma \mathrm{d}t$$

where ν is the outward unit normal to $\partial\Omega$. By carefully analysis, one can show that the integral $\int_{\underline{u}}^{\overline{u}} \int_{\Sigma_t} R_{\Sigma_t}$ is asymptotically nonpositive, and the boundary term is asymptotic to $16\pi(E + a \cdot P)$. To summarize, one obtain

$$E + a \cdot P \geq \frac{1}{16\pi}\int_M \left(\mu + J(\eta) + \frac{\left|\nabla^2 u + k|\nabla u|\right|^2}{|\nabla u|^2}\right) \mathrm{d}\sigma \mathrm{d}t \geq 0.$$

This completes the proof.

4. pp-waves and the equality case $E = |P|$

4.1. *pp-waves*

A pp-wave spacetime (or just a pp-wave for short) is a Lorentzian metric $\mathbf{g}$ defined on $\mathbb{R}^{n+1}$ equipped with a global coordinate chart $\{u, z, x^1, \ldots, x^{n-1}\}$ where

$$\mathbf{g} = 2\mathrm{d}u\mathrm{d}z + S\mathrm{d}z^2 + \sum_{a=1}^{n-1} (\mathrm{d}x^a)^2. \tag{10}$$

Here, the scalar-valued function S depends on z and $x^1, \ldots, x^{n-1}$, but not on u. The vector $\mathbf{Y} := \partial/\partial u$ is a covariantly constant, nowhere vanishing null vector field. Note that a pp-wave is a special case of a null perfect fluid described by (5) with pressure $p = 0$ and carries a global null Killing vector field $\mathbf{Y}$ with $\mathbf{v} = \eta\mathbf{Y}$ for some scalar function η. The term "pp-waves" stands for *plane-fronted* waves (the wave fronts are constant z slices, whose induced metric is flat) with *parallel rays* ($\mathbf{Y}$ is a parallel null vector). These waves can describe situations in which gravitational waves are idealized as propagating along one direction, uniformly across flat wave fronts perpendicular to this direction.

Example 8. The Minkowski metric $\mathbf{g} = -\mathrm{d}t^2 + \sum_{a=1}^n (\mathrm{d}y^a)^2$ is a special case of the pp-wave metric. To see this, we can introduce the null coordinates $U = y^n - t$ and $Z = \frac{1}{2}(y^n + t)$ and re-express the Minkowski metric as $\mathbf{g} = 2\mathrm{d}U\mathrm{d}Z + \sum_{a=1}^{n-1} (\mathrm{d}y^a)^2$. Thus, we recover the pp-wave metric by setting $S = 0$ in (10). Another choice of coordinates is to set

$$U = u - (x^1 + \cdots + x^{n-1}), \quad Z = z, \quad y^a = x^a + z,$$

and express the Minkowski metric as

$$\mathbf{g} = 2\mathrm{d}u\mathrm{d}z + \mathrm{d}z^2 + \sum_{a=1}^{n-1} (\mathrm{d}x^a)^2. \tag{11}$$

Then, we recover the pp-wave metric (10) by setting $S = 1$.

The second expression (11) of the Minkowski metric is more relevant to what we will discuss below. Notice that the induced metric from (11) on the constant u slices is the Euclidean metric with the induced second fundamental form $k \equiv 0$.

We consider the constant u-slices in (10), assuming that $S > 0$ everywhere. Additionally, we use x^n to denote the z coordinate in (10).

Lemma 9. *We assume $S > 0$. Let g and k be respectively the induced metric and second fundamental form on the constant u-slices in (10). Then $(\mathbb{R}^n, g, k)$ is an initial data set where g and k are given by*

$$\begin{aligned} g &= S(\mathrm{d}x^n)^2 + \sum_{a=1}^{n-1}(\mathrm{d}x^a)^2 \\ k_{ab} &= 0 \quad \textit{for } a,b = 1,\dots,n-1 \\ k_{jn} &= \tfrac{1}{2}S^{-\frac{1}{2}}S_{,j} \quad \textit{for } j = 1,\dots,n. \end{aligned}$$

We also have

$$\begin{aligned} \mu &= -\tfrac{1}{2}S^{-1}\Delta' S \\ J &= \frac{1}{2}S^{-\frac{3}{2}}(\Delta' S)\frac{\partial}{\partial x^n}, \end{aligned}$$

and in particular, $\mu = |J|_g$.

Proof. Along a constant u-slice we have $\mathbf{Y} = \partial/\partial u = S^{-\frac{1}{2}}\mathbf{n} + S^{-1}\partial/\partial x^n$ where $\mathbf{n} = \nabla u/|\nabla u|$. To compute the second fundamental form, we use that $\mathcal{L}_{\partial/\partial u}\mathbf{g} = 0$ because $\partial/\partial u$ is Killing and thus

$$\begin{aligned} k_{ij} &= \tfrac{1}{2}\mathcal{L}_{\mathbf{n}}\mathbf{g}(\partial_i,\partial_j) = -\tfrac{1}{2}f^{-1}\mathcal{L}_X\mathbf{g}(\partial_i,\partial_j) = -\tfrac{1}{2}S^{\frac{1}{2}}\mathcal{L}_X g(\partial_i,\partial_j) \\ &= -\tfrac{1}{2}S^{\frac{1}{2}}\left(X_{i,j} + X_{j,i} - 2\Gamma^n_{ij}X_n\right) = S^{\frac{1}{2}}\Gamma^n_{ij}. \end{aligned}$$

Computing the Christoffel symbols gives the desired formula for k. The computations for μ, J are more involved, and we refer to Lemma C.2 of [33]. □

It easily follows that the dominant energy condition holds if and only if $\Delta' S \le 0$. Note that such a function S cannot exist for $n = 3$ because Liouville's theorem says that any superharmonic function on $\mathbb{R}^2$ that is bounded below must be constant. In the next lemma, we summarize the properties that S have such that $(\mathbb{R}^n, g, k)$ is asymptotically flat with the decay rate q.

Lemma 10. *Let $n > 3$. There exist nonconstant smooth functions S on $\mathbb{R}^n$ with the following properties:*

(1) *$\Delta' S \le 0$ everywhere, strictly negative somewhere, and $\Delta' S$ is integrable on $\mathbb{R}^n$.*
(2) *$S \equiv 1$ in $\{|x^n| \ge C\}$ for some constant $C > 0$.*
(3) *$\lim_{\rho\to\infty}\int_{|x'|=\rho} -\sum_{a=1}^{n-1}\partial S/\partial x^a x^a/|x'|\,\mathrm{d}\sigma$ exists and is positive.*
(4) *For each nonnegative integer ℓ and each $\alpha \in (0,1)$, we have $S - 1 \in C^{\ell,\alpha}_{-q}(\mathbb{R}^n)$ with $q = n - 3 - (\ell + \alpha)$.*

Proof. Let F be a smooth nonnegative function on $\mathbb{R}^{n-1}$ with coordinates $x' = (x^1,\dots,x^{n-1})$, such that $F = O(|x'|^{-s})$ for some $s > n-1$. We can solve $\Delta'\psi = -F$ on $\mathbb{R}^{n-1}$ via convolution with the fundamental solution of the Laplacian on Euclidean $\mathbb{R}^{n-1}$. As long as F is not identically zero, $\psi(x')$ will be a positive, globally superharmonic function on $\mathbb{R}^{n-1}$. For $n > 3$, it must have the expansion

$$\psi(x') = A|x'|^{3-n} + (\text{lower order terms}),$$

and since ψ is positive, the constant A must also be positive. Now define

$$S(x', x^n) = 1 + \phi(x^n)\psi(x'),$$

where ϕ is chosen to be any nontrivial, compactly supported, smooth, nonnegative function on $\mathbb{R}$. Note that $\Delta' S = -\phi(x^n)F(x') \le 0$ and is strictly negative somewhere. It is straightforward to verify that S satisfies Items (1), (2), and (3). Since the derivatives of S in the x^n direction do not decay any faster than $|x'|^{3-n}$, we can only conclude that $S-1$ and its derivatives of any order are $O(|x|^{3-n})$. Thus, $S-1 \in C^{\ell,\alpha}_{-q}(\mathbb{R}^n)$ with $q = n-3-\ell-\alpha$ by the definition of weighted Hölder spaces. □

Theorem 11. *For each $n \ge 5$, there exist complete, asymptotically flat initial data sets $(\mathbb{R}^n, g, k)$ that satisfy $\mu = |J|$ and $E = |P| > 0$. These examples have asymptotic decay rate $q > (n-2)/2$.*

Proof. We first describe the case $n > 8$ by Huang and Lee [33]. Choose any S as in Lemma 10 and use this choice in Lemma 9 to construct initial data $(\mathbb{R}^n, g, k)$. We claim that for $n > 8$, this is the desired example. By construction (g, k) is clearly complete. The main task is to show that (g, k) is asymptotically flat at the decay rate $q > (n-2)/2$.

Recall that our asymptotic flatness condition requires $g_{ij} - \delta_{ij} \in C^{2,\alpha}_{-q}(\mathbb{R}^n)$ and $k_{ij} \in C^{1,\alpha}_{-q}(\mathbb{R}^n)$ for some $(n-2)/2 < q < n-2$, and $(\mu, J) \in L^1(\mathbb{R}^n)$. For our (g, k) from Lemma 9, this is equivalent to requiring that $S-1 \in C^{2,\alpha}_{-q}$ for some $(n-2)/2 < q < n-2$ and that $\Delta' S$ is integrable. By Item (4) of Lemma 10, this imposes the condition on n:

$$(n-3) - (2+\alpha) > \frac{n-2}{2} \text{ for some } \alpha \in (0,1), \quad \text{or equivalently,} \quad n > 8.$$

To see that $E = |P| > 0$, we evaluate the ADM energy-momentum by integrating over large capped cylinders. The caps do not contribute, and we can see that

$$E = -P_n = \frac{1}{2(n-1)\omega_{n-1}} \lim_{\rho\to\infty} \int_{|x'|=\rho} -\sum_{a=1}^{n-1} \frac{\partial S}{\partial x^a} \frac{x^a}{|x'|} \, \mathrm{d}\mu > 0,$$

and $P_1 = \cdots = P_{n-1} = 0$.

To extend the example down to dimensions $n \ge 5$, Hirsch and Zhang observed that instead of choosing the constant u-slices in (10) as above, one can choose other spacelike slices, such as graphs over the u-slices. By choosing appropriate graphing functions, they can improve the derivatives in the x^n direction and thus construct examples for $n \ge 5$. See [41]. □

4.2. *Proof that $E = |P|$ implies $E = |P| = 0$*

We review the proof by Huang and Lee [59] for the first statement of the equality theorem.

Theorem 12. *Let (M, g, k) be an n-dimensional asymptotically flat initial data set with decay rate q and assume that the positive mass theorem is true near (g, k). Furthermore, we make the stronger asymptotic assumption that with this decay rate q,*

$$q + \alpha > n-2,$$

where α is the Hölder exponent in the definition of asymptotic flatness. If (g, k) satisfies the dominant energy condition, then for each asymptotically flat end, $E = |P|$ implies that $E = |P| = 0$.

We set up a variational setting to prove the above theorem. Let (M, g, k) be an asymptotically flat initial data set satisfying the dominant energy condition, as well as the assumption $E = |P|$. Given a scalar function f_0 and a vector field X_0 so that $(f_0, X_0) \to (a, b)$, we introduce the *modified Regge–Teitelboim Hamiltonian* $\mathscr{H}$ corresponding to (g, k) and (f_0, X_0) by

$$\mathscr{H}(\gamma, \tau) = (n-1)\omega_{n-1}\left[aE(\gamma,\tau) + b \cdot P(\gamma,\tau)\right] - \int_M \overline{\Phi}|_{(g,k)}(\gamma,\tau) \cdot (f_0, X_0) \, \mathrm{d}\mu_g$$

where the volume measure $\mathrm{d}\mu_g$ and the inner product in the integral are both with respect to g. The functional is obtained from the classical Regge–Teitelboim Hamiltonian in Definition 5 by

replacing the usual constraint operator with the modified constraint operator $\overline{\Phi}|_{(g,k)}$. We also choose the volume form with respect to g.

Define the constraint set $\mathscr{C}_{(g,k)}$ to be the set of asymptotically flat initial data sets (γ,τ) such that $\overline{\Phi}_{(g,k)}(\gamma,k) = \overline{\Phi}|_{(g,k)}(g,k)$. A key property of the modified constraint operator is that if (g,k) satisfies the dominant energy condition, then the dominant energy condition holds for all $(\gamma,\tau) \in \mathscr{C}_{(g,k)}$ close to (g,k). Choosing the pair (f_0, X_0) asymptoting to $(a,b) := (E,-P)$, we apply the positive mass inequality to conclude that (g,k) locally minimizes the functional $\mathscr{H}$ on $\mathscr{C}_{(g,k)}$ as below: For $(\gamma,\tau) \in \mathscr{C}_{(g,k)}$, by the same computation as in (4), we have

$$\mathscr{H}(\gamma,\tau) - \mathscr{H}(g,k) = (n-1)\omega_{n-1}\left[aE(\gamma,\tau) + b\cdot P(\gamma,\tau)\right] - (n-1)\omega_{n-1}\left[aE + b\cdot P\right] \tag{12}$$

$$\geq (n-1)\omega_{n-1}\left[EE(\gamma,\tau) - P\cdot P(\gamma,\tau)\right] \geq 0, \tag{13}$$

where the inequality $E(\gamma,\tau) \geq |P(\gamma,\tau)|$ is used. The equality holds when $(\gamma,\tau) = (g,k)$.

Using the method of Lagrange multipliers, there exists a Lagrange multiplier (f_1, X_1) asymptotic to zero such that

$$D\mathscr{H}|_{(g,k)}(h,w) = \int_M (f_1,X_1)\cdot D\overline{\Phi}|_{(g,k)}(h,w)\,\mathrm{d}\mu_g = \int_M D\overline{\Phi}|^*_{(g,k)}(f_1,X_1)\cdot(h,w)\,\mathrm{d}\mu_g$$

Since the left hand side equals $-\int_M (h,w)\cdot D\overline{\Phi}|^*_{(g,k)}(f_0,X_0)\,\mathrm{d}\mu_g$, just as the first variation of the classical Regge–Teitelboim Hamiltonian (3), we construct a pair $(f,X) = (f_0,X_0) + (f_1,X_1)$ that solves $D\overline{\Phi}|^*_{(g,k)}(f,X) = 0$ and is asymptotic to $(E,-P)$. To summarize, the above construction yields a modified Killing initial data (f,X) with the prescribed asymptotics $(f,X) \to (E,-P)$.

Note that in the non-vacuum case, there need not be a corresponding spacetime Killing vector field. However, intriguingly, the spinor proof by Beig and Chruściel uses the harmonic spinor to construct (f,X) that satisfies a similar over-determined equation with the asymptotic to $(E,-P)$. Their key observation is that for asymptotically flat initial data sets with the optimal decay rate, this implies $E = |P| = 0$. It completes the proof.

4.3. *Proof that $E = |P| = 0$ implies embedding in Minkowski spacetime*

Suppose $E = 0$. We extend the above variational setting to show that the initial data set can be embedded into Minkowski space [39]. Unlike other proofs [17–20, 40], this proof does not rely on the specific method used to establish the $E \geq 0$ theorem. Instead, it characterizes the minimizer of the modified Regge–Teitelboim Hamiltonian by producing additional modified Killing initial data, under the assumption that the positivity of mass holds.

Theorem 13. *Let (M,g,k) be a asymptotically flat initial data set with decay rate $q > (n-2)/2$ and assume that the positive mass theorem is true near (g,k). If (g,k) satisfies the dominant energy condition and $E = 0$, then (M,g,k) is embedded into the Minkowski spacetime.*

As in the proof of Theorem 12, we use the fact that (g,k) minimizes a modified Regge–Teitelboim Hamiltonian $\mathscr{H}$ on the constraint set $\mathscr{C}_{(g,k)}$. Then we invoke the method of Lagrange multiplier to construct a lapse-shift pair (f,X) that solves $D\overline{\Phi}|^*_{(g,k)}(f,X) = 0$ with the prescribed asymptotics $(f,X) \to (a,b)$, where $(a,b) = (E,-P)$.

When $E = |P| = 0$, it turns out that (g,k) minimizes the modified Regge–Teitelboim Hamiltonian for many choices of (a,b), because the last term of (12) is alway zero. By similar computations as in (13), we can choose any a and b with $a = |b|$. This allows us to construct an entire $(n+1)$-dimensional space of lapse-shift pairs that can be thought of as being asymptotic to the $(n+1)$-dimensional space of translational Killing fields on Minkowski space as we approach spatial infinity. Moreover, having so many solutions implies that (g,k) is vacuum, by [33, Corollary 6.6].

Once we know that (g,k) is vacuum, it follows that each of these lapse-shift pairs is actually vacuum Killing initial data for (g,k) by Moncrief's theorem, and we can extend them to become

actual Killing fields on the vacuum development of (g,k). The final step is to show that having such a space of Killing fields on an asymptotically flat Lorentzian manifold that are asymptotic to the $(n+1)$ translational directions implies that the Lorentzian manifold must be Minkowski space.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

The author was partially supported by NSF DMS-2304966. Part of the survey was completed while the author was in residence at the Simons Laufer Mathematical Sciences Institute in Berkeley, California, during the Fall 2024 semester, supported by the NSF DMS-1928930.

References

[1] Y. Choquet-Bruhat, "Beginnings of the Cauchy problem for Einstein's field equations", in *Surveys in Differential Geometry 2015. One Hundred Years of General Relativity*, Surveys in Differential Geometry, vol. 20, International Press, Boston, MA, 2015, p. 1-16.

[2] Y. Choquet-Bruhat, in *General relativity and the Einstein equations*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2009, p. xxvi+785.

[3] Y. Choquet-Bruhat, J. E. Marsden, "Solution of the local mass problem in general relativity", *Commun. Math. Phys.* **51** (1976), no. 3, p. 283-296, http://projecteuclid.org/euclid.cmp/1103900391.

[4] D. R. Brill, S. Deser, "Variational methods and positive energy in general relativity", *Ann. Phys.* **50** (1968), no. 3, p. 548-570.

[5] R. Geroch, "General relativity", in *Differential Geometry (Proceedings of Symposia in Pure Mathematics, Vol. XXVII, Part 2, Stanford University, Stanford, California, 1973)*, American Mathematical Society, Providence, RI, 1975, p. 401-414.

[6] R. Schoen, "Conformal deformation of a Riemannian metric to constant scalar curvature", *J. Differ. Geom.* **20** (1984), no. 2, p. 479-495, http://projecteuclid.org/getRecord?id=euclid.jdg/1214439291.

[7] R. Schoen, S.-T. Yau, "On the proof of the positive mass conjecture in general relativity", *Commun. Math. Phys.* **65** (1979), no. 1, p. 45-76, http://projecteuclid.org/getRecord?id=euclid.cmp/1103904790.

[8] R. Schoen, S.-T. Yau, "The energy and the linear momentum of space-times in general relativity", *Commun. Math. Phys.* **79** (1981), no. 1, p. 47-51, http://projecteuclid.org/getRecord?id=euclid.cmp/1103908887.

[9] R. Schoen, S.-T. Yau, "Complete manifolds with nonnegative scalar curvature and the positive action conjecture in general relativity", *Proc. Nat. Acad. Sci. USA* **76** (1979), no. 3, p. 1024-1025.

[10] R. Schoen, "Variational theory for the total scalar curvature functional for Riemannian metrics and related topics", in *Topics in Calculus of Variations (Montecatini Terme, 1987)*, Lecture Notes in Mathematics, vol. 1365, Springer, 1989, p. 120-154.

[11] J. Lohkamp, "Scalar curvature and hammocks", *Math. Ann.* **313** (1999), no. 3, p. 385-407.

[12] R. Schoen, S.-T. Yau, "Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with nonnegative scalar curvature", *Ann. Math. (2)* **110** (1979), no. 1, p. 127-142.

[13] G. Huisken, T. Ilmanen, "The inverse mean curvature flow and the Riemannian Penrose inequality", *J. Differ. Geom.* **59** (2001), no. 3, p. 353-437, http://projecteuclid.org/getRecord?id=euclid.jdg/1090349447.

[14] Y. Li, "Ricci flow on asymptotically Euclidean manifolds", *Geom. Topol.* **22** (2018), no. 3, p. 1837-1891.

[15] H. L. Bray, D. P. Kazaras, M. A. Khuri, D. L. Stern, "Harmonic functions and the mass of 3-dimensional asymptotically flat Riemannian manifolds", *J. Geom. Anal.* **32** (2022), no. 6, article no. 184.

[16] V. Agostiniani, L. Mazzieri, F. Oronzio, "A Green's function proof of the positive mass theorem", *Commun. Math. Phys.* **405** (2024), no. 2, article no. 54.

[17] R. Schoen, S.-T. Yau, "Proof of the positive mass theorem. II", *Commun. Math. Phys.* **79** (1981), no. 2, p. 231-260.

[18] M. Eichmair, "The Jang equation reduction of the spacetime positive energy theorem in dimensions less than eight", *Commun. Math. Phys.* **319** (2013), no. 3, p. 575-593.

[19] E. Witten, "A new proof of the positive energy theorem", *Commun. Math. Phys.* **80** (1981), no. 3, p. 381-402.
[20] T. Parker, C. H. Taubes, "On Witten's proof of the positive energy theorem", *Commun. Math. Phys.* **84** (1982), no. 2, p. 223-238, http://projecteuclid.org/getRecord?id=euclid.cmp/1103921154.
[21] R. Bartnik, "The mass of an asymptotically flat manifold", *Commun. Pure Appl. Math.* **39** (1986), no. 5, p. 661-693.
[22] L. Ding, "Positive mass theorems for higher dimensional Lorentzian manifolds", *J. Math. Phys.* **49** (2008), no. 2, article no. 022504.
[23] M. Eichmair, L.-H. Huang, D. Lee, R. Schoen, "The spacetime positive mass theorem in dimensions less than eight", *J. Eur. Math. Soc. (JEMS)* **18** (2016), no. 1, p. 83-121.
[24] S. Hirsch, D. Kazaras, M. Khuri, "Spacetime harmonic functions and the mass of 3-dimensional asymptotically flat initial data for the Einstein equations", *J. Differ. Geom.* **122** (2022), no. 2, p. 223-258.
[25] J. Lohkamp, "Skin Structures on Minimal Hypersurfaces", Preprint, 2015, https://arxiv.org/abs/1512.08249.
[26] J. Lohkamp, "Hyperbolic Geometry and Potential Theory on Minimal Hypersurfaces", Preprint, 2015, https://arxiv.org/abs/1512.08251.
[27] J. Lohkamp, "Skin Structures in Scalar Curvature Geometry", Preprint, 2015, [math.DG], https://arxiv.org/abs/1512.08252.
[28] J. Lohkamp, "The Higher Dimensional Positive Mass Theorem II", Preprint, 2016, [math.DG], https://arxiv.org/abs/1612.07505.
[29] R. Schoen, S.-T. Yau, "Positive scalar curvature and minimal hypersurface singularities", in *Surveys in Differential Geometry 2019. Differential Geometry, Calabi-Yau Theory, and General Relativity. Part 2*, Surveys in Differential Geometry, vol. 24, Internatial Press, Boston, MA, 2022, p. 441-480.
[30] N. Smale, "Generic regularity of homologically area minimizing hypersurfaces in eight-dimensional manifolds", *Commun. Anal. Geom.* **1** (1993), no. 2, p. 217-228.
[31] O. Chodosh, S. F. Mantoulidis, Christos, "Generic regularity for minimizing hypersurfaces in dimensions 9 and 10", Preprint, 2023, [math.DG], https://arxiv.org/abs/2302.02253.
[32] W. W.-Y. Wong, "A positive mass theorem for two spatial dimensions", Preprint, [gr-qc], https://arxiv.org/abs/1202.6279.
[33] L.-H. Huang, D. A. Lee, "Bartnik mass minimizing initial data sets and improvability of the dominant energy scalar", *J. Differ. Geom.* **126** (2024), no. 2, p. 741-800.
[34] A. Ashtekar, G. T. Horowitz, "Energy-momentum of isolated systems cannot be null", *Phys. Lett. A* **89** (1982), no. 4, p. 181-184.
[35] P. F. Yip, "A strictly-positive mass theorem", *Commun. Math. Phys.* **108** (1987), no. 4, p. 653-665, http://projecteuclid.org/euclid.cmp/1104116629.
[36] R. Beig, P. T. Chruściel, "Killing vectors in asymptotically flat space-times. I. Asymptotically translational Killing vectors and the rigid positive energy theorem", *J. Math. Phys.* **37** (1996), no. 4, p. 1939-1961.
[37] P. T. Chruściel, D. Maerten, "Killing vectors in asymptotically flat space-times. II. Asymptotically translational Killing vectors and the rigid positive energy theorem in higher dimensions", *J. Math. Phys.* **47** (2006), no. 2, article no. 022502.
[38] R. Bartnik, "Phase space for the Einstein equations", *Commun. Anal. Geom.* **13** (2005), no. 5, p. 845-885, http://projecteuclid.org/euclid.cag/1144438298.
[39] L.-H. Huang, D. A. Lee, "Equality in the spacetime positive mass theorem II", Preprint, 2023, [math.DG], https://arxiv.org/abs/2302.06040.
[40] S. Hirsch, Y. Zhang, "The case of equality for the spacetime positive mass theorem", *J. Geom. Anal.* **33** (2023), no. 1, article no. 30.
[41] S. Hirsch, Y. Zhang, "Initial data sets with vanishing mass are contained in pp-wave spacetimes", Preprint, 2024, [gr-qc], https://arxiv.org/abs/2403.15984.
[42] H. L. Bray, W. P. Minicozzi II, M. Eichmair, L.-H. Huang, S.-T. Yau, K. Uhlenbeck, R. Kusner, F. Codá Marques, C. Mese, A. Fraser, "The mathematics of Richard Schoen", *Not. Amer. Math. Soc.* **65** (2018), no. 11, p. 1349-1376.
[43] L.-H. Huang, D. A. Lee, "Trapped surfaces, topology of black holes, and the positive mass theorem", *Not. Amer. Math. Soc.* **69** (2022), no. 4, p. 536-545.
[44] D. A. Lee, in *Geometric Relativity*, Graduate Studies in Mathematics, vol. 201, American Mathematical Society, Providence, RI, 2019, xii+361 pages.
[45] X. Wang, "The mass of asymptotically hyperbolic manifolds", *J. Differ. Geom.* **57** (2001), no. 2, p. 273-299, http://projecteuclid.org/euclid.jdg/1090348112.
[46] P. T. Chruściel, M. Herzlich, "The mass of asymptotically hyperbolic Riemannian manifolds", *Pacific J. Math.* **212** (2003), no. 2, p. 231-264.
[47] L. Andersson, M. Cai, G. J. Galloway, "Rigidity and positivity of mass for asymptotically hyperbolic manifolds", *Ann. Henri Poincaré* **9** (2008), no. 1, p. 1-33.
[48] P. T. Chruściel, G. J. Galloway, L. Nguyen, T.-T. Paetz, "On the mass aspect function and positive energy theorems for asymptotically hyperbolic manifolds", *Class. Quantum Grav.* **35** (2018), article no. 115015.

[49] L.-H. Huang, H. C. Jang, D. Martin, "Mass rigidity for hyperbolic manifolds", *Commun. Math. Phys.* **376** (2020), no. 3, p. 2329-2349.

[50] A. Sakovich, "The Jang equation and the positive mass theorem in the asymptotically hyperbolic setting", *Commun. Math. Phys.* **386** (2021), no. 2, p. 903-973.

[51] J. Corvino, L.-H. Huang, "Localized deformation for initial data sets with the dominant energy condition", *Calc. Var. Partial Differ. Equations* **59** (2020), no. 1, article no. 42.

[52] P. T. Chruściel, "On the invariant mass conjecture in general relativity", *Commun. Math. Phys.* **120** (1988), no. 2, p. 233-248, http://projecteuclid.org/euclid.cmp/1104177748.

[53] T. Regge, C. Teitelboim, "Role of surface integrals in the Hamiltonian formulation of general relativity", *Ann. Phys.* **88** (1974), p. 286-318.

[54] V. Moncrief, "Spacetime symmetries and linearization stability of the Einstein equations. I", *J. Math. Phys.* **16** (1975), p. 493-498.

[55] M. Eichmair, "The size of isoperimetric surfaces in 3-manifolds and a rigidity result for the upper hemisphere", *Proc. Amer. Math. Soc.* **137** (2009), no. 8, p. 2733-2740.

[56] D. Christodoulou, N. Ó. Murchadha, "The boost problem in general relativity", *Commun. Math. Phys.* **80** (1981), no. 2, p. 271-300, http://projecteuclid.org/euclid.cmp/1103919879.

[57] P. T. Chruściel, "On the energy of the gravitational field at spatial infinity", in *Conference on Mathematical Relativity (Canberra, 1988)*, Proceedings of the Centre for Mathematical Analysis, Australian National University, vol. 19, 1989.

[58] H. Bray, S. Hirsch, D. Kazaras, M. Khuri, Y. Zhang, "Spacetime harmonic functions and applications to mass", in *Perspectives in Scalar Curvature*, vol. 2, World Scientific Publishing, Hackensack, NJ, 2023, p. 593-639.

[59] L.-H. Huang, D. A. Lee, "Equality in the spacetime positive mass theorem", *Commun. Math. Phys.* **376** (2020), no. 3, p. 2379-2407.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 29-52
https://doi.org/10.5802/crmeca.275

Review article / *Article de synthèse*

Gluing for the Einstein constraint equations

Recollement pour les équations de contrainte d'Einstein

Justin Corvino[a]

[a] Lafayette College, Department of Mathematics, Easton, PA 18042, USA
E-mail: corvinoj@lafayette.edu

Abstract. Initial data for the Einstein equations must satisfy a system of nonlinear partial differential equations, the Einstein constraint equations. Constructing interesting solutions of the constraint equations can then lead to interesting spacetime evolutions. Over the past twenty-five years, various gluing methods to construct solutions of the constraints have shed light on some long-standing questions. In this article we survey some of the methods and results achieved to date.

Résumé. Les données initiales des équations d'Einstein doivent satisfaire un système d'équations aux dérivées partielles non linéaires, les équations de contrainte d'Einstein. La construction de solutions intéressantes des équations de contrainte peut alors conduire à des évolutions intéressantes de l'espace-temps. Au cours des vingt-cinq dernières années, diverses méthodes de recollement permettant de construire des solutions des équations de contrainte ont permis de faire la lumière sur certaines questions de longue date. Dans cet article, nous passons en revue quelques-unes des méthodes et des résultats obtenus à ce jour.

Keywords. Einstein constraint equations, Geometric analysis, Gluing constructions.

Mots-clés. Equations de contrainte d' Einstein, Analyse géométrique, Méthodes de recollement.

Funding. AMS-Simons Research Enhancement Grant for PUI Faculty.

Manuscript received 7 November 2024, accepted 19 November 2024.

1. Introduction and preliminaries

The foundational work of Yvonne Choquet-Bruhat[1] firmly established that the Einstein field equation can be treated as an evolution problem [5]. The initial data for the vacuum equation are given by a Riemannian manifold (M,g) along with a symmetric $(0,2)$-tensor K, that determine a vacuum spacetime, i.e. a time-orientable Ricci-flat Lorentzian manifold $(\mathscr{S},\bar{g})$ in which M is a Cauchy surface embedded with first and second fundamental forms g and K, which play the role of the initial metric and its time derivative (in appropriate coordinates). In the non-vacuum case, the initial data will be augmented with data for whatever physical fields are being modeled.

The initial data g and K are not freely prescribed, but must adhere to the Gauss and Codazzi equations relating the ambient geometry to the fundamental forms. The Einstein equation involves the Ricci tensor, and so from appropriate traces of the Gauss and Codazzi equations along with the Einstein equation, one deduces a coupled, nonlinear *underdetermined* elliptic system for g and K, the exploration of which has provided a rich playground for mathematical physicists and geometric analysts, with deep connections to topology and geometry, in part forged by advances in the theory of partial differential equations. Robust methods to construct

[1] For her exposition on this, see her book [1], the classic survey articles [2,3], as well as [4] for her comments on some of the history of the problem.

solutions of the constraint equations allow for constructions, theoretical and numerical, of spacetimes. Understanding the solution space to the constraint equations is thus intimately tied to an understanding of the solution space of the Einstein field equation, and a parametrization of the solution space of the constraints would ostensibly isolate the true gravitational degrees of freedom [6], cf. [7].

A fulsome account of the work on the constraint equations, which one can trace back at least eighty years to classical work of Lichnerowicz [8], is well beyond the scope of this article. Rather, we focus on some relatively recent work of the past twenty-five years on constructing solutions of the Einstein constraint equations by *gluing* techniques, discussing some of the methods and applications without giving an exhaustive account. Gluing techniques can be used to address the following natural question: given multiple solutions of the constraint equations, is there another solution (on a connected manifold) which contains chosen regions from the given initial data sets (at least approximately)? There is of course the related question about spacetimes: given regions of multiple spacetimes solving Einstein's field equation, is there a (connected) spacetime satisfying the field equation that contains them?

From the initial data point of view, if one can construct an initial data set which contains regions obtained from several spacetimes, the resulting evolution of the initial data will contain regions isometric to ones in the given spacetimes, by finite speed of propagation. That said, one might naturally try to glue spacetime regions together directly, or at least utilize the Lorentzian structure in the construction. In this direction we mention two different recent series of works. Hintz constructs spacetimes with multiple small black holes by a gluing construction using geometric singular analysis (pseudodifferential b-calculus) techniques (see [9], cf. [10]). In a different direction, Aretakis, Czimek and Rodnianski achieve a suite of gluing results by *characteristic gluing*, which leverages the *null structure* of spacetime, as surveyed with additional references in [11]. While these works are interesting, we do not have the time in the present short survey article to do justice to either, and we will focus on constructing solutions of the constraint equations on a spacelike hypersurface (though we do mention some compelling results obtained by the methods of [11] in Section 4 below).

1.1. *Overview of some spacelike gluing results*

A natural interpretation of the above aim of gluing is that of constructing initial data for a relativistic N-body problem [12], including, say, multi-black hole spacetimes. Such constructions have been achieved in earlier works [13–15], while more recent works have used gluing to construct more exotic such initial data sets, such as the N-body configurations with gravitational shielding from [16], multi-localized initial data sets [17], as well as initial data with small black holes [10]. Naturally, the issue of the two-body and N-body problems in general relativity dates back a century (see [8, 18, 19] and references therein) and various sorts of approaches have been taken to produce initial data to model such spacetimes.

Connected sum construction of initial data was achieved by Isenberg, Mazzeo and Pollack in [20] by the conformal method, and Chruściel and Delay [21] observed how to combine conformal and localized gluing techniques (introduced in [22, 23]) to perform connected sum constructions which leave the data unperturbed outside a connecting neck region (cf. [24, 25] for related results, and [26, 27] for analogous surgery results). Of particular note, [28] established that there is no topological obstruction to admitting an asymptotically Euclidean solution of the vacuum Einstein constraint equations: for a closed manifold Σ and $p \in \Sigma$, there is a vacuum asymptotically flat solution of the Einstein constraints on $\Sigma \setminus \{p\}$. An intriguing implementation of the conformal gluing method was employed by Stavrov in [29] to construct families of initial data to model point particle limits in general relativity (cf. recent work of Hintz [9, 10] along these lines).

Gluing constructions have also shed light on the asymptotics of solutions of the constraints. The original motivation for the construction in [22] was to understand to what extent the asymptotics determine the interior of a vacuum initial data set, in particular how special are Schwarzchild asymptotics. That initial work showed that any compact subset of any time-symmetric asymptotically flat vacuum initial data set is also contained in a vacuum time-symmetric initial data set which is *exactly* Schwarzschild near infinity. While that in itself may seem intriguing, one immediately turns to the spacetime evolution of such data. To be precise, the question open at the time was whether there exist any nontrivial asymptotically simple vacuum spacetimes (sometimes called *purely radiative* spacetimes), a question answered in the affirmative in [30, 31] by constructing suitable families of initial data which are Schwarzschild with small mass in a *uniform* neighborhood of infinity, evolving to an asymptotically hyperboloidal hypersurface, and then invoking the work of [32] for the asymptotics of the evolution.

Another question that has been addressed by gluing methods is that of *gravitational shielding*. A remarkable result was obtained by Carlotto and Schoen [16], who showed that any region of any vacuum initial data set can be localized within a conical region in a vacuum initial data set which is *flat* outside the conical region. While Bartnik's parabolic quasi-spherical construction [33] produced examples of nontrivial asymptotically flat vacuum initial data (time-symmetric, so scalar-flat) containing a flat region, their localized gluing construction produced striking examples with non-compact flat regions. The results in [16] were proven in spirit along the lines of [21–23], with a number additional of technical obstacles to surmount. While the works [21–23] established that any bounded region in an asymptotially flat vacuum solution can be realized as a subdomain in a solution in which each asymptotic end is a slice of a Schwarzschild or Kerr spacetime, the method of proof requires that the end be tuned appropriately to the given data (in terms of the asymptotically conserved quantities: ADM energy-momentum, center-of-mass and angular momentum), whereas the regions in the result of Carlotto and Schoen are glued to flat exteriors outside the conical regions. In any case, all these results illustrate that unique continuation fails spectacularly for the vacuum constraints.

In this article we overview aspects of gluing by the conformal method [20] and compare with the localized approach. While there are many excellent references for the conformal method, localized gluing constructions are relatively newer, and are still being developed, so we will focus on some details of the approach to localized gluing that dates back to the works [22, 23]. In the spirit of "something old and something new", we illustrate the localized gluing technique by applying it to the time-symmetric Einstein–Maxwell constraint equations.

1.2. *Preliminaries*

1.2.1. *The Einstein constraint equations*

The vacuum constraint equations arise from the following. Consider a spacetime splitting $(\mathscr{S} = I \times M, \bar{g})$, with I an interval and (M, g) a Riemannian hypersurface given as a level set of the coordinate x^0 on I (so $\partial/\partial x^0$ is timelike). With $G = \mathrm{Ric}(\bar{g}) - (1/2)R(\bar{g})\bar{g}$ the Einstein tensor, $G_{\alpha\beta}n^\beta$ does not contain any second time derivatives $\bar{g}_{\mu\nu,00}$ of the metric components, where n is a timelike unit normal to M. In fact, using the Gauss and Codazzi equations, one obtains, with $G(n,\cdot)$ a one-form along M,

$$G(n,n) = \tfrac{1}{2}(R(g) - |K|_g^2 + (\mathrm{tr}_g K)^2) \tag{1}$$

$$G(n,\cdot) = \mathrm{div}_g(K - (\mathrm{tr}_g K)g). \tag{2}$$

If we consider a cosmological constant Λ, and let $G_\Lambda = G + \Lambda\bar{g}$, and if T is the stress-energy tensor, the Einstein equation takes the form $G_\Lambda = \kappa T$ for some constant $\kappa > 0$. (In spacetime dimension four, $\kappa = 8\pi G/c^4$.) We observe that $\mu := T(n,n)$ is the energy density as measured by

an observer with four-velocity cn, and $c^{-1}J := -c^{-1}T(n,\cdot)$ is corresponding momentum density one-form. The dominant energy condition that the energy-momentum vector of the matter fields is causal future-pointing is given by $\mu \geq |J|_g$. Combining the Einstein equation with (1)–(2) yields the Einstein constraint equations

$$\tfrac{1}{2}(R(g) - |K|_g^2 + (\mathrm{tr}_g K)^2) = \kappa\mu + \Lambda \tag{3}$$

$$\mathrm{div}_g(K - (\mathrm{tr}_g K)g) = \kappa J. \tag{4}$$

In the vacuum ($T = 0$) time-symmetric ($K = 0$) case, the system reduces to the constant scalar curvature (CSC) condition $R(g) = 2\Lambda$. Unless noted, we take $\Lambda = 0$, in which case the time-symmetric vacuum constraints reduce to $R(g) = 0$.

For a non-vacuum example, consider initial data for the time-symmetric Einstein–Maxwell equations, often given as a Riemannian metric g and an electric field E. The constraint (3) becomes $R(g) = 2\kappa|E|_g^2$. In any source-free region, we must also have $\mathrm{div}_g E = 0$. For simplicity we will treat E as a vector field and study the map $\Psi(g,E) = (R(g) - 2\kappa|E|_g^2, \mathrm{div}_g E)$. We note that it could be more natural to specify the field through a differential form η_E, to which a metric g would associate a vector field E_g in such as way that the condition $\mathrm{div}_g E_g = 0$ would follow *for any* g in a source-free region. We will see the echo of this in the gluing method below. For further exploration of specifying the initial data for fields, see [34].

We will focus on the constraints in dimension $n = 3$, though much of what is discussed below extends readily to $n \geq 3$. We employ the Einstein summation convention of summing over repeated upper and lower indices; a semicolon denotes a covariant derivative, and a comma denotes a partial derivative.

1.2.2. *Some classical solutions*

For later use, we recall a family of time-symmetric solutions to the Einstein–Maxwell constraint equations, each obtained from data on a spacelike hypersurface in a Reissner–Nordström spacetime: $g_{\mathrm{RN}} = (1 - (2m/r) + (Q^2/r^2))^{-1}\,\mathrm{d}r^2 + r^2\mathring{g}_{\mathbb{S}^2}$, with $E_{\mathrm{RN}} = (Q/r^2)\sqrt{1-(2m/r)+(Q^2/r^2)}(\partial/\partial r) =: (Q/r^2)e_r$, where $\mathring{g}_{\mathbb{S}^2}$ is the unit round metric. Since $R(g_{\mathrm{RN}}) = 2Q^2/r^4$, we see the units are such that $\kappa = 2$. We will just employ the data on an asymptotic end $r = |x| \gg 1$, and so we do not pose any relation between m and Q, though that could affect the Reissner–Nordström spacetime structure.

For $Q = 0$, the Reissner–Nordström metric reduces to the Schwarzschild metric, which has vanishing scalar curvature. We write the metric in isotropic form, and we can also re-center the coordinates to obtain $(g_S^{m,c})_{ij}(x) = (1 + m/(2|x-c|))^4\delta_{ij}$. A metric $g_{ij} = u^4\delta_{ij}$ which is conformally Euclidean has vanishing scalar curvature precisely when u is Euclidean harmonic. If we require that u tends to 1 at infinity, to impose asymptotically Euclidean geometry, then u is either identically 1 or it must have singularities, such as for the Schwarzschild metric with $m \neq 0$, and for the following family of solutions.

As recalled earlier, it has been a longstanding question as to how to construct solutions to the Einstein equations to model the interaction of multiple bodies. A classical solution of the vacuum constraints is given by vacuum Brill–Lindquist metrics (see [35,36], cf. [8]): given a finite set $\mathfrak{p} = \{p_1,\dots,p_N\}$ of N points in $\mathbb{R}^3$ and $\mathfrak{m} = (m_1,\dots,m_N) \in \mathbb{R}^N$, we consider the metric

$$(g_{\mathrm{BL}}^{\mathfrak{m},\mathfrak{p}})_{ij}(x) = \left(1 + \sum_{k=1}^{N} \frac{m_k}{2|x-p_k|}\right)^4 \delta_{ij} \tag{5}$$

for $x \in \mathbb{R}^3 \setminus \mathfrak{p}$. This metric is conformally Euclidean with harmonic conformal factor, and thus has vanishing scalar curvature, providing a time-symmetric vacuum initial data set. If we take all $m_k > 0$, then the data has $N+1$ asymptotically flat ends. By an ingenious use of the method of images, Misner [37] modified the conformal factor, incorporating an infinite sequence of

inversions, to produce a scalar-flat metric with two asymptotically flat ends and N Einstein–Rosen bridges, hence obtaining a type of gluing construction by conformal means. In another direction, one can modify the Brill–Lindquist data by localized gluing techniques to produce data on $\mathbb{R}^3$, even allowing $\mathfrak{p}$ to be infinite; see [17] along with comments in Section 4 below.

1.2.3. *The conformal method*

The vacuum Einstein constraints form an *underdetermined* elliptic system for g and K. The conformal method specifies some parts of g and K, leaving a determined elliptic system. This approach can be traced back to[2] Lichnerowicz [8], Choquet-Bruhat [38], York and Ó Murchadha [6,39,40]. The formulation below is especially well suited to the constant mean curvature (CMC) case with $\mathrm{tr}_g K$ constant[3] (for a variation see [40]). It is based on the transverse-traceless (TT) decomposition of symmetric two-tensors. In this approach, a conformal class $\mathscr{C}$ of metrics is given, say $\mathscr{C} = [\gamma]$. Furthermore a scalar function τ and a $(0,2)$-tensor σ are given on M^n, where σ is TT with respect to γ: $\mathrm{div}_\gamma \sigma = 0$, $\mathrm{tr}_\gamma \sigma = 0$.

From this data we seek a scalar function $u > 0$ and a vector field W such that, with

$$g = u^{\frac{4}{n-2}}\gamma, \quad K = u^{-2}(\sigma + L_\gamma W) + \frac{\tau}{n} u^{\frac{4}{n-2}}\gamma,$$

where $(L_\gamma W)_{ij} = W_{i;j} + W_{j;i} - (2/n)(\mathrm{div}_\gamma W)\gamma_{ij}$ is the conformal Killing operator, (g, K) satisfies the vacuum constraint equations, which take the form

$$\Delta_\gamma u - \frac{n-2}{4(n-1)} R(\gamma) u + \frac{n-2}{4(n-1)} |\sigma + L_\gamma W|^2_\gamma u^{-\frac{3n-2}{n-2}} - \frac{n-2}{4n}\tau^2 u^{\frac{n+2}{n-2}} = 0 \tag{6}$$

$$\mathrm{div}_\gamma(L_\gamma W) - \frac{n-1}{n} u^{\frac{2n}{n-2}}\, \mathrm{d}\tau = 0. \tag{7}$$

The system is a semilinear elliptic system for (u, W). In the CMC case (constant τ), the system decouples: the second equation becomes $\mathrm{div}_\gamma(L_\gamma W) = 0$. On a closed manifold, or asymptotically flat manifold with W decaying suitably near infinity, multiplying by W and integrating by parts gives $L_\gamma W = 0$, i.e. W is a *conformal Killing field*. In this case W does not affect K or the equation for u, and we can just take $W = 0$ (which is the only option in the absence of conformal symmetries). Then the conformal method comes down to solving the *Lichnerowicz equation*

$$\Delta_\gamma u - \frac{n-2}{4(n-1)} R(\gamma) u + \frac{n-2}{4(n-1)} |\sigma|^2_\gamma u^{-\frac{3n-2}{n-2}} - \frac{n-2}{4n}\tau^2 u^{\frac{n+2}{n-2}} = 0.$$

As noted earlier, we will focus on $n = 3$, in which case the above becomes

$$\Delta_\gamma u - \tfrac{1}{8} R(\gamma) u + \tfrac{1}{8}|\sigma|^2_\gamma u^{-7} - \tfrac{1}{12}\tau^2 u^5 = 0. \tag{8}$$

1.2.4. *The conformal method and asymptotics*

When gluing initial data with the conformal method (see Section 2), the resulting data can be made arbitrarily close to the given data outside the gluing region. A similar scenario arises when using conformal perturbations to arrange conformally Euclidean, or more generally *harmonic*, asymptotics. Indeed Schoen and Yau [42] showed that a scalar-flat asymptotically flat metric can be approximated by one which is conformally Euclidean outside a compact set, and so that the ADM energy is perturbed by an arbitrarily small amount. The method of proof is straightforward: along a large annular region, patch the original metric to a Euclidean metric near infinity; for large enough annuli, we get approximately scalar-flat data. One then employs a conformal factor to reimpose the vacuum constraint, with appropriate estimates on the size of the perturbation of the metric and the ADM energy. Such an approximation result is useful for proving the Positive

[2] One can find an early analysis of the conformally Euclidean case in [18], which also includes some early study of static spaces and asymptotically Euclidean spaces; see [4] for more comments on the history.

[3] See [41], cf. [7,34], for discussion of the conformal method outside the vacuum CMC regime.

Energy Theorem [43] and the Riemannian Penrose Inequality, since it simplifies the asymptotics with only a small change to the energy, and Bray termed such metrics *harmonically flat* at infinity [44].

An analogous non-time-symmetric version established in [23], *harmonic asymptotics*, has also proved useful in this and other contexts [45, 46]. The formulation is a modification of the conformal method above so that the principal part of the constraints operator in the asymptotic region is a diagonal Laplacian. The asymptotic form of the data in harmonic asymptotics is

$$g_{ij} = u^4\delta_{ij}, \quad K_{ij} = u^2\left(X_{i,j} + X_{j,i} - \tfrac{1}{2}X^k_{,k}\delta_{ij}\right). \tag{9}$$

This form is usually written in terms of the momentum tensor $\pi = K - (\mathrm{tr}_g K)g$, in which case this becomes $\pi_{ij} = u^2(X_{i,j} + X_{j,i} - X^k_{,k}\delta_{ij})$.

The proof for harmonic asymptotic approximation of asymptotically flat initial data sets is similar to the time-symmetric case. Given asymptotically flat data (g, K), smoothly patch the metric to Euclidean while cutting off K to zero, over a large annular region $R < |x| < 2R$, with the resulting data $(\hat{g}, \hat{\pi})$ (a family of data depending on R). Let $(\mathscr{L}_g X)_{ij} = X_{i;j} + X_{j;i} - (\mathrm{div}_g X)g_{ij}$. One seeks to reimpose the constraints and obtain harmonic asymptotics by considering data of the form $\bar{g} = u^4\hat{g}$ and $\bar{\pi} = u^2(\hat{\pi} + \mathscr{L}_{\hat{g}}X)$, or $\bar{K} = u^2(\hat{K} + \mathscr{L}_{\hat{g}}X + (1/2)(\mathrm{div}_{\hat{g}}X)\hat{g})$. Observe that it is not the conformal Killing operator $L_{\hat{g}}X$ that appears here, in contrast to the standard formulation of the conformal method. In any case, $(\hat{g}, \hat{K})$ is an approximate solution of the vacuum constraints (better for larger R), and one seeks to perturb to an exact solution, for which one considers the map

$$(u, X) \mapsto T(u, X) = \left(R(\bar{g}) - |\bar{K}|^2_{\bar{g}} + (\mathrm{tr}_{\bar{g}}\bar{K})^2, \mathrm{div}_{\bar{g}}(\bar{K} - (\mathrm{tr}_{\bar{g}}K)\bar{g})\right).$$

Outside a compact set, i.e. for $|x| > 2R$, the vacuum constraints $T(u, X) = 0$ can be written in Cartesian components as (omitting metric subscripts for the Euclidean metric)

$$8\Delta u = u\left(-|\mathscr{L}X|^2 + \tfrac{1}{2}(\mathrm{tr}(\mathscr{L}X))^2\right), \quad \Delta X^i = -4u^{-1}u_{,j}(\mathscr{L}X)^j_i + 2u^{-1}u_{,i}\mathrm{tr}(\mathscr{L}X). \tag{10}$$

Solutions in spaces with suitable decay for $u - 1$ and X will admit partial expansions of the form $u(x) = 1 + (a/|x|) + O(|x|^{-2})$, and $X^i(x) = (b^i/|x|) + O(|x|^{-2})$. Using (9), one finds the ADM energy is $2a$, and the linear momentum is $P^i = -(1/2)b^i$.

The linearization of $T(u, X)$ at $(1, 0)$ is readily found to be Fredholm of index zero between appropriate spaces [23], of which (10) is certainly indicative. The possible existence of a finite-dimensional cokernel is accommodated by finding a complementary finite-dimensional space of *compactly supported* tensors, and showing that one can find (h, k) near $(0, 0)$ in this complementing space along with (u, X) close to $(1, 0)$ (for large R), such that $(\bar{g} + h, \bar{K} + k)$ solves the vacuum constraints. It can also be shown that the energy and linear momentum can be perturbed an arbitrarily small amount in this construction [23].

In a certain sense, these approximation constructions establishing that a kind of asymptotic behavior is suitably dense can be construed as a type of gluing construction, gluing a given interior to an exterior end in a certain family. Good approximate solutions form a family parametrized by R, which can be corrected with suitable perturbations. In this case the family of exterior data, either the harmonically flat metrics, or more generally data with harmonic asymptotics, is a large family, as compared to a data like the Schwarzschild family, parametrized by m and c. Just as in the conformal gluing we recall below, the given data is perturbed a small (for large R) amount.

Harmonic asymptotics yield a particularly convenient form of the asymptotics for the energy and momenta. For some other applications, one might want to make controlled construction of solutions of the constraints with more general asymptotics. For instance, recent work of Fang, Szeftel and Touati [47, 48] constructs suitable classes of vacuum initial data compatible with the stability results for Minkowski spacetime and for black hole stability. The basic approach

is to start with an approximate solution with a desired asymptotic behavior and then perturb to a solution of the constraints, with estimates on the perturbation which ensure the desired decay rate. The perturbation is as above, utilizing a conformal change of metric, together with a perturbation of the momentum tensor of the form $\mathscr{L}_\gamma X$ (for a suitable metric γ), plus a complementing compactly supported perturbation.

1.2.5. *Remarks on localized deformations*

A fundamental issue for constructing solutions to the constraints by gluing methods is how to obtain controlled perturbations of the constraints map (where $\pi = K - (\mathrm{tr}_g K)g$)

$$(g,\pi) \mapsto \Phi(g,\pi) := (R(g) - |K|_g^2 + (\mathrm{tr}_g K)^2, \mathrm{div}_g \pi). \tag{11}$$

As recalled above, modifying data with a suitable conformal transformation has proven fruitful. Since a conformal factor u solves an elliptic equation, we do not expect to be able to effectively contain the support of $u-1$. With the linearization stability results of Fischer and Marsden [49,50] in mind, one can naturally move outside the conformal regime and consider a general form of perturbations, modifying (g,π) to $(g+\delta g, \pi+\delta\pi)$. The vacuum constraints then give an underdetermined elliptic system for the perturbations $(\delta g, \delta\pi)$. This freedom is what has been employed in [22, 23] and numerous subsequent works to achieve more direct control on the support of the deformation tensors.

Of course, the analysis will focus on the linearized constraints operator, which morally speaking should be locally surjective when the formal adjoint of the linearization has trivial kernel. The presence of nontrivial kernel indicates symmetries in the spacetime evolution of the data, as in [51], so that except in special situations, we expect to be able to effectively perturb the constraints operator. That said, in many gluing constructions, the approximate solution approaches a special configuration which admits symmetries, and this requires some extra care to handle.

To make this precise, we consider the linearization and formal L^2-adjoint for the scalar curvature and vacuum constraints operators. The well-known formulas for the linearization of the scalar curvature operator and its adjoint are given by

$$DR_g(h) = -\Delta_g(\mathrm{tr}_g h) + \mathrm{div}_g \mathrm{div}_g h - h \cdot_g \mathrm{Ric}(g), \quad DR_g^*(f) = -(\Delta_g f)g + \mathrm{Hess}_g f - f\mathrm{Ric}(g). \tag{12}$$

At a Ricci-flat metric, an element of the kernel of DR_g^* is precisely a function f with vanishing Hessian, which in the Euclidean case means f is a constant plus a linear combination of the Cartesian coordinate functions x^1, x^2, x^3.

While the formulas for the linearization of Φ and its formal adjoint are straightforward to compute in the general case, they simplify tremendously at the flat data, about which we will be perturbing. Let $\mathring{g}$ be the Euclidean metric on $\mathbb{R}^3$. If we let $D\mathring{\Phi} = D\Phi_{(\mathring{g},0)}$, we find from (11)

$$D\mathring{\Phi}(h,\omega) = (DR_{\mathring{g}}(h), \mathrm{div}_{\mathring{g}}\omega), \quad D\mathring{\Phi}^*(f,Z) = \left(DR_{\mathring{g}}^*(f), -\tfrac{1}{2}L_Z\mathring{g}\right),$$

where $(L_Z g)_{ij} = Z_{i;j} + Z_{j;i}$ is the Lie derivative, so that at $\mathring{g}$ in Cartesian coordinates, we have $(L_Z\mathring{g})_{ij} = Z_{i,j} + Z_{j,i}$. Now, $L_Z g = 0$ precisely for a Killing vector field Z, which at $\mathring{g}$ must be a linear combination of the generators of translations and rotations. Thus the kernel of $D\mathring{\Phi}^*$ gives a ten-dimensional space of potential obstructions to localized gluing constructions for the vacuum Einstein equations, near Minkowskian initial data. To accommodate for this, we glue to model families of solutions that are governed by ten parameters which can cover this approximate cokernel, such as the Kerr family of asymptotically flat initial data (see, e.g. [21, Appendix F]). We will see in Section 3 below how the mass and center-of-mass parameters can be used to handle the kernel of $DR_{\mathring{g}}^*$, while in the general case, the linear and angular momentum parameters correspond to the translational and rotational Killing fields [21,23]. In general the kernel elements of $D\Phi^*_{(g,\pi)}$ are related to spacetime symmetries [51], and are called *Killing Initial Data*.

2. Gluing by the conformal method

We now discuss the connected sum construction of Isenberg, Mazzeo and Pollack [20]. We start with solutions of the vacuum constraints on manifolds M_1 and M_2 with constant mean curvature τ, given on each as a Riemannian metric γ_j, and a TT tensor μ_j with respect to γ_j, so that the second fundamental form is given $K_j = \mu_j + (\tau/3)\gamma_j$ (recall that we consider dimension three).

To construct a connected sum $M = M_1 \sharp M_2$ of disjoint manifolds M_1 and M_2, one removes a ball around points $p_1 \in M_1$ and $p_2 \in M_2$ and identifies the union of the complements along the boundary spheres (the construction could also be applied where p_1 and p_2 are points on the same manifold, corresponding to adding a handle). We seek to construct a family of solutions on $M_1 \sharp M_2$, such that for any compactly contained domains away from p_1 and p_2, there are solutions in the family arbitrarily close to the original data. The approximate solutions are obtained using conformal blowups of $M_1 \setminus \{p_1\}$ and $M_2 \setminus \{p_2\}$, producing solutions with asymptotically cylindrical ends. The connected sum is then performed by cutting off far down on each cylinder and making appropriate identifications. We assume that K_j is nontrivial, and we also make the nondegeneracy assumption that there are no nontrivial conformal Killing vector fields on M_j that vanish at p_j.

In normal coordinate neighborhoods $U_j \ni p_j$ around each point, the metric takes the form $\gamma_j = \mathrm{d}r_j^2 + r_j^2 h_j(r_j)$, where h_j gives a family of metrics on $\mathbb{S}^2$, with $h_j(0) = \mathring{g}_{\mathbb{S}^2}$ the unit round metric. For $j = 1,2$, let ψ_j be a positive smooth function which is identically one on most of M_j, and which is $r_j^{1/2}$ near each p_j, say for $r_j < \mathring{r}$, and with $\psi_j = 1$ outside of $r_j < 2\mathring{r}$ (take $\log(\mathring{r}^{-2}) > 1$). With the change in radial coordinate $t_j = -\log r_j$, we find $\psi_j^{-4}\gamma_j = \mathrm{d}t_j^2 + \mathring{g}_{\mathbb{S}^2} + O(\mathrm{e}^{-2t_j})$ [25]. One introduces a large parameter $T \gg -2\log \mathring{r} > 1$ to specify how far along each neck we go before transitioning smoothly to an *exactly* cylindrical metric $\mathring{\gamma} = \mathrm{d}s_j^2 + \mathring{g}_{\mathbb{S}^2}$. Indeed with coordinate $s_j = -\log(r_j/\mathring{r}) - (T/2)$, we have $\psi_j^{-4}\gamma_j = \mathrm{d}s_j^2 + \mathring{g}_{\mathbb{S}^2} + \mathrm{e}^{-T}\mathrm{e}^{-2s_j}\mathring{r}^2\hat{h}_j$, and we can smoothly transition in the region $(-1,-1/2)\times\mathbb{S}^2$ to the exactly cylindrical metric on $(-1/2,\infty)\times\mathbb{S}^2$. We appropriately form a quotient along the cylindrical pieces $(-1/2,1/2)\times\mathbb{S}^2$ where the change of coordinates identifies s_1 with $-s_2$; note that $s_j = 0$ where $r_j = \mathring{r}\mathrm{e}^{-T/2}$, while $s_j = -T/2$ where $r_j = \mathring{r}$. With $s = s_1$ for $s \le 0$ and $s = -s_2$ for $s \ge 0$, then on $C_T \cong [-T/2, T/2]\times\mathbb{S}^2$, the metric is $\gamma_T = \mathrm{d}s^2 + \mathring{g}_{\mathbb{S}^2} + \mathrm{e}^{-T}\cosh(2s)\hat{h}_T$, where $\hat{h}_T$ and its covariant derivatives with respect to the cylindrical metric are bounded.

We have now put a metric on the connected sum, patching together the conformal rescalings. Observe that $\psi_j^2\mu_j$ is TT for $\psi_j^{-4}\gamma$, and just like for γ_T, we can patch together the conformally rescaled TT-tensors along the cylinder, to obtain an approximate solution μ_T. We want to conformally rescale appropriately to get a suitable approximate solution to the constraint equations. To do this, we patch the conformal factors $\psi_T := \chi_{1,T}\psi_1 + \chi_{2,T}\psi_2$, in such a way that ψ_T is the sum $\psi_1 + \psi_2$ inside the region corresponding to $[-(T/2)+1, (T/2)-1]\times\mathbb{S}^2$, i.e. most of C_T. In this region, then

$$\psi_T = r_1^{1/2} + r_2^{1/2} = (\mathring{r}\mathrm{e}^{-s-\frac{T}{2}})^{1/2} + (\mathring{r}\mathrm{e}^{s-\frac{T}{2}})^{1/2} = 2\mathring{r}^{1/2}\mathrm{e}^{-\frac{T}{4}}\cosh\left(\frac{s}{2}\right).$$

In the exactly cylindrical piece $-1/2 < s < 1/2$, $\psi_T^4\gamma_T$ is readily seen to have scalar curvature *zero*. By rotational symmetry, then, it is *exactly Schwarzschild* in this region, with mass $2\mathring{r}\mathrm{e}^{-T/2}$. Thus for T large, we have a connected sum with the middle of the neck being a neighborhood of the very small horizon sphere.

From here, one can first modify μ_T to a γ_T-TT tensor $\tilde{\mu}_T = \mu_T - \sigma_T$, with the perturbation σ_T which is globally small, with norm exponentially decaying in T [20, Proposition 5]. This step employs the conformal Killing operator, and this is where the nondegeneracy assumption is used. The proof is interesting and nontrivial, but we focus on the Lichnerowicz operator here for the sake of brevity, noting its role in constant scalar curvature gluing (see, e.g. [25]).

Having sorted out $\tilde{\mu}_T$, one then solves the Lichnerowicz equation

$$\mathscr{N}_T(u) := \Delta_{\gamma_T} u - \tfrac{1}{8} R(\gamma_T) u + \tfrac{1}{8} |\tilde{\mu}_T|^2_{\gamma_T} u^{-7} - \tfrac{1}{12}\tau^2 u^5 = 0.$$

We make some remarks on the proof.

First, ψ_T can be used as an approximate solution, since one can bound $\|\mathscr{N}_T(\psi_T)\|_{C^{k,\alpha}} \leq Ce^{-T/2}$, where C is independent of T large [20, Proposition 6], cf. [25, Proposition 4.6]. Indeed outside of C_T, $\mathscr{N}_T(\psi_T) = 0$. Now, where $\psi_T = \psi_1 + \psi_2$, we have $\Delta_{\mathring{\gamma}}\psi_T - (1/8)R(\mathring{\gamma})\psi_T = 0$ (we note $R(\mathring{\gamma}) = 2$ for the cylindrical metric). As γ_T is approximately cylindrical in $[-1,1]\times\mathbb{S}^2$, with precise error estimates, and with corresponding estimates on $\tilde{\mu}_T$ and ψ_T, we can estimate $\mathscr{N}_T(\psi_T)$ in this region. The other regions in C_T are handled likewise, using that in $[-T/2,-1]\times\mathbb{S}^2$ we have $\mathscr{N}_T(\psi_1) = 0$, and in $[1, T/2]\times\mathbb{S}^2$ we have $\mathscr{N}_T(\psi_2) = 0$.

The next step is to get uniform (in T) bounds on the inverse of the linearization $\mathscr{L}_T = D\mathscr{N}_T|_{\psi_T}$. We observe

$$\mathscr{L}_T = \Delta_{\gamma_T} - \tfrac{1}{8}\left(R(\gamma_T) + 7|\tilde{\mu}_T|^2_{\gamma_T}\psi_T^{-8} + \tfrac{10}{3}\tau^2\psi_T^4\right). \tag{13}$$

Showing that $\mathscr{L}_T$ is invertible for all T large and bounding the inverse are both handled by blowup arguments. Indeed, suppose the existence of a sequence $T_m \nearrow \infty$, with $\sup_M |\eta_m| = 1$ but $\mathscr{L}_{T_m}\eta_m = 0$. By the Schauder estimate and reindexing a subsequence, we can assume that η_m converges to η in C^2. There are two basic scenarios: either we can also arrange the subsequence so that $\sup|\eta_m| \geq c > 0$ on $M_1 \setminus \{p_1\}$ or $M_2 \setminus \{p_2\}$ (note that $(M_1\sharp M_2, \gamma_T)$ contains more of $(M_1,\gamma_1)\cup(M_2,\gamma_2)$ isometrically as T grows), or η_m converges locally uniformly to zero on these two pieces and the point where $\sup_M|\eta_m| = 1$ is attained travels further down the cylinder. By translation we can center the supremum point, and obtain a nontrivial bounded η on the cylinder $\mathbb{R}\times\mathbb{S}^2$, satisfying $\Delta_{\mathring{\gamma}}\eta - (1/4)\eta = 0$; this follows from (13) by the estimates of $\tilde{\mu}_T$ and ψ_T. As there are no nontrivial such η (by separation of variables or the maximum principle), we have a contradiction.

Thus we must be in the former case, say η is nontrivial on $M_1 \setminus \{p_1\}$. In this case, γ_T converges to $\psi_1^{-4}\gamma_1$, $|\tilde{\mu}_T|^2_{\gamma_T}$ converges to $\psi_1^{12}|\mu_1|^2_{\gamma_1}$, so that $\mathscr{L}_T$ (13) converges to

$$\tilde{\mathscr{L}} = \Delta_{\psi_1^{-4}\gamma_1} - \tfrac{1}{8}\left(R(\psi_1^{-4}\gamma_1) + 7|\mu_1|^2_{\gamma_1}\psi_1^4 + \tfrac{10}{3}\tau^2\psi_1^4\right).$$

The conformal Laplacian satisfies a covariance property,

$$\Delta_{\psi_1^{-4}\gamma_1}\eta - \tfrac{1}{8}R(\psi_1^{-4}\gamma_1)\eta = \psi_1^5\left(\Delta_{\gamma_1}(\psi_1^{-1}\eta) - \tfrac{1}{8}R(\gamma_1)\psi_1^{-1}\eta\right),$$

while since $(\gamma_1, K_1 = \mu_1 + (\tau/3)\gamma_1)$ solves the vacuum constraints, we have $R(\gamma_1) = |\mu_1|^2_{\gamma_1} - (2/3)\tau^2$, and so we see that $\tilde{\mathscr{L}}\eta = 0$ can be re-written

$$\left(\Delta_{\gamma_1} - |\mu_1|^2_{\gamma_1} - \tfrac{1}{3}\tau^2\right)(\psi_1^{-1}\eta) = 0.$$

The operator $(\Delta_{\gamma_1} - |\mu_1|^2_{\gamma_1} - (1/3)\tau^2)$ has trivial kernel for nontrivial K_1, which we assume. Thus it admits a positive Green's function G which has a pole or order r_1^{-1} at p_1, with r_1 the distance to p_1. Since $|\eta| \leq 1$ and $|\psi^{-1}| \leq r_1^{-1/2}$, $\psi^{-1}\eta$ must extend smoothly across p_1, and so we conclude it must vanish, contradicting that η is nontrivial on $M_1 \setminus \{p_1\}$.

Similar arguments can be used to control the inverse [20, Proposition 8] (cf. [25, 29]). Writing a Taylor expansion $\mathscr{N}_T(\psi_T + \eta) = \mathscr{N}_T(\psi_T) + \mathscr{L}_T(\eta) + \mathscr{Q}_T(\eta)$, one can then show that for large T, the following map is a contraction on a suitable ball around 0 with small T-dependent radius in a Hölder space:

$$\eta \mapsto -\mathscr{L}_T^{-1}(\mathscr{N}_T(\psi_T) + \mathscr{Q}_T(\eta)) = -\mathscr{L}_T^{-1}(N_T(\psi_T + \eta) - \mathscr{L}_T(\eta)).$$

The keys are that ψ_T is a sufficiently good approximate solution, and there is suitable control on $\mathscr{L}_T^{-1}$ uniform in T large.

The fixed point η_T furnishes the desired conformal factor $\tilde{\psi}_T = \psi_T + \eta_T$, and the solution $(\tilde{\psi}_T^4 \gamma_T, \tilde{\psi}_T^{-2}\tilde{\mu}_T + (\tau/3)\tilde{\psi}_T^4 \gamma_T)$ of the constraints on $M_1 \sharp M_2$. The difference between the solution and the starting data decays exponentially in T on compact subsets of each $M_j \setminus \{p_j\}$.

This result gives us a way to combine solutions to the vacuum constraints (or to add a handle to a solution), with the constituent parts approximately represented in the final solution. There are analogous results in the asymptotically Euclidean and asymptotically hyperboloidal setting, and an interesting implementation [29] of conformal gluing (including a different family of approximate solutions) to construct initial data with some features of point particle solutions.

A natural question to ask is whether and under what conditions one can preserve compact subdomains in the original data. Results in this direction, see e.g. [21, 24, 25], use localized deformations, to which we now turn.

3. Localized gluing

In this section we will outline steps to glue an asymptotically flat initial data set across a large annulus to an initial data set suitably chosen from a given family of initial data. The perturbation will be localized to the annular region.

To bring the ideas across, we will present a proof sketch of the following. For any $R > 0$, let $\mathscr{E}_R = \{x \in \mathbb{R}^3 : |x| > R\}$, and let $\mathscr{E} = \mathscr{E}_1$.

Theorem. *Let (g, E) be asymptoticaly flat, time-symmetric source-free initial data for the Einstein–Maxwell equations on $\mathscr{E}$, with ADM mass $m(g) \neq 0$. For sufficiently large R, there are source-free initial data $(\bar{g}, \bar{E})$ on $\mathscr{E}$ which agree with the given data data on $\mathscr{E} \setminus \mathscr{E}_R$, and on $\mathscr{E}_{2R}$ agree with asymptotically flat data from a spacelike slice in a suitable Reissner–Nordström spacetime, with the same electric charge.*

The statement is for an asymptotic end (which could be outside nontrivial charge distributions), whereas if (M, g, E) comprised a nontrivial complete asymptotically flat solution of the time-symmetric Einstein–Maxwell constraints, then since $R(g) = 2\kappa|E|_g^2 \geq 0$, we would have $m(g) > 0$, by the Positive Mass Theorem.

The corresponding theorem for the non-time symmetric case requires the exterior slice to be chosen suitably in the Kerr–Newman family. For simplicity of exposition, we focus on the time-symmetric case here. Before we begin, we note that non-vacuum gluing results using the conformal method have been explored in [52].

We remark that the argument will show the construction works for *interior* gluing, i.e., the asymptotic exterior could be taken from the given data, while the interior would then be from a suitable member of the model family. For example, in the time-symmetric vacuum case, the interior can be made to be a member of the Schwarzschild family. If this interior solution has positive mass, then the resulting data will be complete with two asymptotically flat ends and will contain the Schwarzschild minimal sphere; as the gluing region will be far out in the asymptotic end, this will be an outermost minimal surface in the end, so the Penrose inequality restricts the mass based on that of the exterior. In any case, we note that interior gluing has been used for long-time existence results for the evolution problem [53, 54].

3.1. *Proof of the Theorem*

The method of proof we follow here is the same as in [22, 23], cf. [21]. We smoothly patch together the given data to data from a slice in the model spacetime across an annular region $\mathscr{A}_R = \{x : R < |x| < 2R\}$ to form an approximate solution. As R increases, the metric approaches $\mathring{g}$,

where $\mathring{g}_{ij} = \delta_{ij}$ in the asymptotic coordinates, and E tends to zero (in the general case, K also tends to zero). As such, while the approximation improves with increasing R, it approaches a solution which admits nontrivial kernel of the formal adjoint of linearization of the constraints. This finite-dimensional kernel, arising from symmetries, is precisely the obstruction to a local (and localized) perturbation result, as we now outline.

The first key to achieving localized perturbation in our framework is a weighted elliptic estimate (17), which in the time-symmetric case derives from an absolute estimate (16) without boundary terms. The estimates are used to guarantee that the solution obtained lies in an appropriate weighted space, so that it extends smoothly across $\mathscr{A}_R$. Like the conformal method above, we need to get uniform estimates on a family of linearized operators, but this can only be done transverse to the finite-dimensional approximate kernel of the adjoint of linearized operator (see estimate (18) below). The presence of this approximate cokernel for large R leads us to employ a Lyapunov–Schmidt reduction: we use the weighted linear elliptic estimates and a Picard-type quasi-Newtonian iteration to solve a nonlinear projected problem that puts the constraint data into a finite-dimensional space. We handle the obstruction space by choosing a suitable member of the model family, in this case Reissner–Nordström.

To be precise, we recall the following operator introduced earlier:

$$\Psi(g,E) = (R(g) - 2\kappa|E|_g^2, \operatorname{div}_g E).$$

For a perturbative analysis, we compute its linearization and formal L^2-adjoint, for which one readily finds (and recall (12))

$$D\Psi_{(g,E)}(h,X) = \left(DR_g(h) - 4\kappa g(E,X) - 2\kappa h(E,E), \operatorname{div}_g(X) + \tfrac{1}{2}d(\operatorname{tr}_g h)(E)\right),$$

and hence

$$D\Psi^*_{(g,E)}(f,\psi) = \left(DR_g^*(f) - 2\kappa f E^\flat \otimes E^\flat - \tfrac{1}{2}\operatorname{div}_g(\psi E)g, -\operatorname{grad}_g \psi - 4\kappa f E\right). \tag{14}$$

These simplify at the flat data, about which we will be perturbing. Let $\mathring{g}$ be the Euclidean metric on $\mathbb{R}^3$, with connection $\mathring{\nabla}$. If we let $D\mathring{\Psi} = D\Psi_{(\mathring{g},0)}$ with formal L^2-adjoint $D\mathring{\Psi}^*$, then

$$D\mathring{\Psi}(h,X) = (DR_{\mathring{g}}(h), \operatorname{div}_{\mathring{g}} X), \quad D\mathring{\Psi}^*(f,\psi) = (DR_{\mathring{g}}^*(f), -\operatorname{grad}_{\mathring{g}}\psi).$$

The obstruction space is given by the elements (f,ψ) in the kernel of $D\mathring{\Psi}^*$. As observed earlier, then f must be a constant plus a linear combination of the Cartesian coordinate functions x^1, x^2, x^3, while ψ must be a constant. The end of the proof will show how to cover these directions by picking appropriately the parameters (mass, center of mass and charge) of the Reissner–Nordström data to which we glue, as we will detail in Section 3.1.7.

3.1.1. *Weighted spaces*

We suppose $(\overline{\Omega}, \mathring{g})$ is compact, connected with smooth nonempty boundary $\partial\Omega$. It suffices for our purpose here to consider Ω to be an annular region. Given $N > 0$, and $0 < r_1 < r_0$, we let $\tilde{\rho} : \mathbb{R} \to \mathbb{R}$ be a smooth nondecreasing function with $\tilde{\rho}(t) = 0$ for $t \le 0$, $\tilde{\rho}(t) = \mathrm{e}^{-N/t}$ for $0 < t < r_1$, and $\tilde{\rho}(t) = 1$ for $t \ge r_0$. We define ρ on Ω by $\rho(x) = \tilde{\rho}(\mathrm{d}(x,\partial\Omega))$, where $\mathrm{d}(x,\partial\Omega)$ is the $\mathring{g}$-distance of x to $\partial\Omega$; we choose r_0 suitably small, so that in particular ρ is smooth. On an annular region, ρ decays to zero at the boundary spheres, and is identically 1 near the middle of the annulus.

Let $\|u\|^2_{L^2_\rho(\Omega)} = \int_\Omega |u|^2 \rho \, \mathrm{d}\mu_{\mathring{g}}$, while for k a nonnegative integer, let $\|u\|^2_{H^k_\rho(\Omega)} = \sum_{\ell=0}^k \|\mathring{\nabla}^\ell u\|_{L^2_\rho(\Omega)}$. The resulting Hilbert spaces can in fact can be identified as the closure of $C^\infty(\overline{\Omega})$ in the relevant

norm [23, Lemma 2.1]. Furthermore by [55, Proposition 2.10], upon choosing N suitably large, there is a constant $C > 0$ such that for all $u \in H^k_\rho(\Omega)$,

$$\|u\rho^{1/2}\|_{H^k(\Omega)} \le C\|u\|_{H^k_\rho(\Omega)}. \tag{15}$$

As such, for a bounded sequence in $H^k_\rho(\Omega)$, there is a $u \in H^k_\rho(\Omega)$ and a subsequence u_i for which $u_i\rho^{1/2}$ converges to $u\rho^{1/2}$ in $H^{k-1}(\Omega)$ and weakly in $H^k(\Omega)$, while u_i converges weakly to u in $H^k_\rho(\Omega)$. We note that for metrics g in a $C^k(\overline{\Omega})$-bounded set, the norms defined as above but using the measures $\mathrm{d}\mu_g$ and connection ∇_g are uniformly equivalent.

The linearized operators enjoy estimates (16)–(17) in these weighted spaces, which are used in solving the linearized equation variationally (see Section 3.1.4). There are also weighted Hölder spaces built with the appropriate scaling to capture the interior Schauder estimates (22)–(23) that follow for the solutions of the linearized equations. These spaces are defined with a natural scaling compatible with the weight function ρ, and in fact there is a suite of weighted spaces compatible with different weights ρ. In [22], we first dealt with power decay, so $\rho = d^N$ near $\partial\Omega$, which has a simple weighted estimate. While the exponential weighting was discussed there, the weighted Hölder spaces used by Chruściel and Delay [21] give a more elegant formulation of the interior estimates, and allow for various kinds of weights all in one framework.

We follow an equivalent implementation as in [55]. We take a smooth function $0 < \phi < 1$ with $\phi = d^2$ near the boundary, and with $\overline{B_{\phi(x)}(x)} \subset \Omega$ for all $x \in \Omega$. We let φ be of the form $\varphi = \phi^r\rho^s$, with r and s real numbers, and let

$$\|u\|_{C^{k,\alpha}_{\phi,\varphi}(\Omega)} = \sup_{x\in\Omega}\left(\sum_{j=0}^{k}\varphi(x)\phi^j(x)\|\mathring{\nabla}^j u\|_{C^0(B_{\phi(x)}(x))} + \varphi(x)\phi^{k+\alpha}(x)[\mathring{\nabla}^k u]_{0,\alpha;B_{\phi(x)}(x)}\right).$$

Every derivative is matched with a power of ϕ, corresponding to $\tilde{\rho}'(t) = Nt^{-2}\tilde{\rho}(t)$, and by iterating this we have $|\phi^k\rho^{-1}\nabla^k\rho|$ is bounded.

3.1.2. *Integral estimates*

The key to time-symmetric estimates is that DR^*_g is overdetermined-elliptic. In fact

$$\mathrm{Hess}_g f = DR^*_g(f) - \frac{1}{n-1}(\mathrm{tr}_g(DR^*_g(f)) + fR(g))g + f\mathrm{Ric}(g),$$

from which we immediately see that there is a constant $C > 0$ such that on any open set Ω, $\|f\|_{H^2(\Omega)} \le C(\|DR^*_g(f,\psi)\|_{L^2(\Omega)} + \|f\|_{H^1(\Omega)})$.

With an application of Rellich's lemma we can replace the right-most norm by $\|f\|_{L^2(\Omega)}$, and together with (14) we obtain

$$\|f\|_{H^2(\Omega)} + \|\psi\|_{H^1(\Omega)} \le C(\|D\Psi^*_{(g,E)}(f,\psi)\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)} + \|\psi\|_{L^2(\Omega)}). \tag{16}$$

We can replace the above norms with the weighted Sobolev norms by a method used in [22], namely we can replace Ω with $\Omega_\varepsilon = \{x \in \Omega : d(x,\partial\Omega) > \varepsilon\}$ for small $\varepsilon > 0$, multiply the (squared) estimates by $\tilde{\rho}'(\varepsilon) \ge 0$ and integrate:

$$\int_0^{\varepsilon_0}\tilde{\rho}'(\varepsilon)(\|f\|^2_{H^2(\Omega_\varepsilon)} + \|\psi\|^2_{H^1(\Omega_\varepsilon)})\,\mathrm{d}\varepsilon \le C\int_0^{\varepsilon_0}\tilde{\rho}'(\varepsilon)(\|D\Psi^*_{(g,E)}(f,\psi)\|^2_{L^2(\Omega_\varepsilon)} + \|f\|^2_{L^2(\Omega_\varepsilon)} + \|\psi\|^2_{L^2(\Omega_\varepsilon)})\,\mathrm{d}\varepsilon.$$

Integration by parts together with the coarea formula lead to

$$\|(f,\psi)\|_{H^{2,1}_\rho(\Omega)} := \|f\|_{H^2_\rho(\Omega)} + \|\psi\|_{H^1_\rho(\Omega)} \le C(\|D\Psi^*_{(g,E)}(f,\psi)\|_{L^2_\rho(\Omega)} + \|f\|_{L^2_\rho(\Omega)} + \|\psi\|_{L^2_\rho(\Omega)}). \tag{17}$$

We note that (16), and hence (17), holds for *any* data (g,E), with a constant uniform for (g,E) in a bounded $C^2(\overline{\Omega}) \times C^1(\overline{\Omega})$ neighborhood.

While we focus on the time-symmetric case, we do note that the estimate

$$\|Z\|_{H^1_\rho(\Omega)} \le C(\|L_Z g\|_{L^2_\rho(\Omega)} + \|Z\|_{L^2_\rho(\Omega)})$$

for Z is more subtle. The proof from [55, Lemma 5.1] is a variation on the argument from [23, Lemma 4.1]; both use the elementary inequality $\Delta_{\mathring{g}}\rho \ge (1/2)N^2 d^{-4}\rho$ near $\partial\Omega$ and an integration-by-parts argument, and the estimate holds for N sufficiently large. The analogue of (17) for the full constraints operator follows.

3.1.3. *Uniform estimate transverse to cokernel*

We can promote (17) to a coercivity estimate, at least in directions transverse to the cokernel $\mathscr{K}$ at flat data. Recall $\mathscr{K} = \mathrm{span}\{1, x^1, x^2, x^3\} \oplus \mathrm{span}\{1\} \subset H^2_\rho(\Omega) \times H^1_\rho(\Omega) =: H^{2,1}_\rho(\Omega)$. Fix a nontrivial smooth bump function ζ supported on Ω. Let S_g be the $L^2(\Omega, \mathrm{d}\mu_g)$-orthogonal complement of $\zeta\mathscr{K}$.

We claim there is a $C > 0$ and a $C^2(\overline{\Omega}) \times C^1(\overline{\Omega})$ neighborhood $\mathscr{U}$ of the flat data $(\mathring{g}, 0)$ such that for all data $(g, E) \in \mathscr{U}$ and all $(f, \psi) \in S_g$,

$$\|(f,\psi)\|_{H^{2,1}_\rho(\Omega)} \le C\|D\Psi^*_{(g,E)}(f,\psi)\|_{L^2_\rho(\Omega)}. \tag{18}$$

The same holds if S_g is replaced by any closed subspace S of $H^{2,1}_\rho(\Omega)$ with $S \cap \mathscr{K} = \{0\}$.

There are a couple ways one can prove this. For instance, one can argue along the lines of the proof of [25, Proposition 3.1], establishing the *unweighted* analogue of (18) on Ω_ε, with a uniform constant for ε small, and proceed as in the proof of (17). Another approach is to use (17) along with a standard compactness argument: from a sequence $(f_i, \psi_i) \in S_{g_i}$, with $(g_i, E_i) \to (\mathring{g}, 0)$, $\|(f_i,\psi_i)\|_{H^{2,1}_\rho(\Omega)} = 1$ and $\|D\Psi^*_{(g,E)}(f_i,\psi_i)\|_{L^2_\rho(\Omega)} \to 0$, one can construct a nontrivial $(f,\psi) \in S_{\mathring{g}} \cap \mathscr{K}$, which furnishes a contradiction. Indeed as noted after (15), there is a $(f,\psi) \in H^{2,1}_\rho(\Omega)$ such that, upon re-indexing a subsequence, $(f_i,\psi_i)\rho^{1/2}$ converges to $(f,\psi)\rho^{1/2}$ in $H^1(\Omega) \times L^2(\Omega)$. In particular, (f_i,ψ_i) converges in $L^2_{\mathrm{loc}}(\Omega)$ to (f,ψ), so that $D\mathring{\Psi}^*(f,\psi) = 0$ weakly. Hence $(f,\psi) \in \mathscr{K}$. Since (f_i,ψ_i) is $L^2(\mathrm{d}\mu_{g_i})$-orthogonal to $\zeta\mathscr{K}$, $(f,\psi) \in S_{\mathring{g}}$. Moreover, since $(f_i,\psi_i) \to (f,\psi)$ in $L^2_\rho(\Omega)$, we see from (17) that (f,ψ) is nontrivial, which is a contradiction.

3.1.4. *The projected problem*

With ζ, $\mathscr{K}$ and S_g as above, let Π_g be the $L^2(\Omega, \mathrm{d}\mu_g)$-orthogonal projection to S_g, and let $\mathring{\Pi} = \Pi_{\mathring{g}}$. For simplicity, we may arrange that ζ is supported where $\rho = 1$. We study the map $(g, E) \mapsto \mathring{\Pi}(\Psi(g, E))$, with linearization $(h, X) \mapsto \mathring{\Pi}(D\Psi_{(g,E)}(h, X))$.

The goal is to solve $\Psi(g, E) = 0$. We start by solving a projected problem:

$$\mathring{\Pi}(\Psi(g_0 + h, E_0 + X)) = 0,$$

where (g_0, E_0) is close to $(\mathring{g}, 0)$. The linearization of this is

$$\mathring{\Pi}(D\Psi_{(g_0,E_0)}(h, X)) = -\mathring{\Pi}(\Psi(g_0, E_0)). \tag{19}$$

To solve this we consider the functional on $H^{2,1}_\rho(\Omega) \cap S_{g_0}$ given by

$$\mathscr{G}_0(f,\psi) = \int_\Omega \left(\frac{1}{2}\rho|D\Psi^*_{(g_0,E_0)}(f,\psi)|^2_{g_0} + \Pi_{g_0}(\Psi(g_0,E_0)) \cdot (f,\psi)\right) \mathrm{d}\mu_{g_0}.$$

If $\mathscr{G}_0$ is stationary at (f_0, ψ_0), the Euler–Lagrange equation

$$\Pi_{g_0}(D\Psi_{(g_0,E_0)}\rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0)) = -\Pi_{g_0}(\Psi(g_0,E_0))$$

is implied by the following, which holds for all (u, v) compactly supported:

$$\int_\Omega (\Pi_{g_0}(D\Psi_{(g_0,E_0)}\rho D\Psi^*_{(g_0,E_0)}(f,\psi)) + \Pi_{g_0}(\Psi(g_0,E_0))) \cdot (u,v)\, \mathrm{d}\mu_{g_0}.$$

This follows trivially for $(u,v) \in \zeta\mathscr{K}$, while for $(u,v) \in S_{g_0}$, the leading projection operator in the above integral equation can be removed, and so it follows by stationarity of the functional at $(f,\psi) \in H^{2,1}_\rho(\Omega) \cap S_{g_0}$.

We conclude that $\Psi(g_0,E_0) + D\Psi_{(g_0,E_0)}\rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0) \in \zeta\mathscr{K}$, and so we see that (19) is solved with $(h,X) = \rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0)$. Moreover, combining (18) with $\mathscr{G}_0(f_0,\psi_0) \le 0$ we find

$$\|(f_0,\psi_0)\|_{H^{2,1}_\rho(\Omega)} \le C\|\Pi_{g_0}(\Psi(g_0,E_0))\|_{L^2_{\rho^{-1}}(\Omega)} \le C'\|\mathring{\Pi}(\Psi(g_0,E_0))\|_{L^2_{\rho^{-1}}(\Omega)}. \tag{20}$$

3.1.5. *Pointwise estimates*

The above shows we can solve the desired equation at the linear level, with an integral estimate. To iterate linear corrections to solve the nonlinear problem, we require a pointwise estimate of the variational solution, in part to establish the convergence of a quasi-Newtonian iteration scheme.

The operator $(f,\psi) \mapsto D\Psi_{(g,E)}D\Psi^*_{(g,E)}(f,\psi)$, and in fact for our purposes, the weighted version

$$(f,\psi) \mapsto \rho^{-1}D\Psi_{(g,E)}\rho D\Psi^*_{(g,E)}(f,\psi) =: (L_{11}f + L_{12}\psi, L_{21}f + L_{22}\psi), \tag{21}$$

is an elliptic operator of mixed order that admits Douglis–Nirenberg weights [56]: L_{ij} is order $s_i + t_j$, where $s_1 = 0$, $s_2 = -1$, $t_1 = 4$, $t_2 = 3$, i.e. L_{11} is order four, L_{12} and L_{21} are order three, and L_{22} is order two. For example, at the flat data $(\mathring{g},0)$, $D\mathring{\Psi}D\mathring{\Psi}^*(f,\psi) = (2\Delta^2_{\mathring{g}}f, -\Delta_{\mathring{g}}\psi)$. The diagonal terms of (21) have the analogous principal parts, but there are also off-diagonal terms (though they are small for E small), and terms from the weight, which can in principle blow up on approach to the boundary at a rate which is a power of d^{-1}. This behavior can be accommodated in the interior elliptic estimates by a scaling argument, which is efficiently coded by the weighted Hölder norms introduced earlier. Indeed if we write L_{ij} in local coordinates as $L_{ij} = \sum_{|\beta|\le s_i+t_j} b^\beta_{ij}\partial^\beta$, then it is not hard to see that for $0 < \alpha < 1$, there is a $C > 0$ such that $\|b^\beta_{ij}\|_{C^{-s_i,\alpha}_{\phi,\phi^{s_i+t_j-|\beta|}}(\Omega)} \le C$. We record the following interior Schauder estimate (see [55, Theorem 5.8], e.g.), with $\varphi_j = \phi^{(3/2)+4-t_j}\rho^{1/2}$, i.e. $\varphi_1 = \phi^{3/2}\rho^{1/2}$ and $\varphi_2 = \phi^{5/2}\rho^{1/2}$:

$$\begin{aligned}&\|f\|_{C^{4,\alpha}_{\phi,\varphi_1}(\Omega)} + \|\psi\|_{C^{3,\alpha}_{\phi,\varphi_2}(\Omega)}\\ &\quad\le C(\|L_{11}f + L_{12}\psi\|_{C^{0,\alpha}_{\phi,\phi^4\varphi_1}(\Omega)} + \|L_{21}f + L_{22}\psi\|_{C^{1,\alpha}_{\phi,\phi^2\varphi_2}(\Omega)} + \|f\|_{L^2_\rho(\Omega)} + \|\psi\|_{L^2_{\phi^2\rho}(\Omega)}).\end{aligned} \tag{22}$$

We apply this to get a pointwise estimate of (f_0,ψ_0), where $(h,X) = \rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0)$ is the variational solution to (19). The integral terms can be estimated by (20). We can also replace terms in (22) involving the operator by the their projected values: since $\rho = 1$ on the support of ζ,

$$\rho^{-1}D\Psi_{(g_0,E_0)}\rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0) - \rho^{-1}\mathring{\Pi}(D\Psi_{(g_0,E_0)}\rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0)) \in \zeta\mathscr{K}.$$

Using the fact that $\zeta\mathscr{K}$ is finite-dimensional (so all norms on it are equivalent) and an elementary compactness argument, one can derive (cf. the proof of [55, Proposition 6.4])

$$\begin{aligned}&\|f_0\|_{C^{4,\alpha}_{\phi,\varphi_1}(\Omega)} + \|\psi_0\|_{C^{3,\alpha}_{\phi,\varphi_2}(\Omega)} + \|(f_0,\psi_0)\|_{H^{2,1}_\rho(\Omega)}\\ &\quad\le C(\|\rho^{-1}\mathring{\Pi}(\Psi(g_0,E_0))\|_{C^{0,\alpha}_{\phi,\phi^4\varphi_1}(\Omega)\times C^{1,\alpha}_{\phi,\phi^2\varphi_2}(\Omega)} + \|f_0\|_{L^2_\rho(\Omega)} + \|\psi_0\|_{L^2_{\phi^2\rho}(\Omega)})\\ &\quad\le C(\|\mathring{\Pi}(\Psi(g_0,E_0))\|_{C^{0,\alpha}_{\phi,\phi^4\varphi_1\rho^{-1}}(\Omega)\times C^{1,\alpha}_{\phi,\phi^2\varphi_2\rho^{-1}}(\Omega)} + \|\mathring{\Pi}(\Psi(g_0,E_0))\|_{L^2_{\rho^{-1}}(\Omega)})\end{aligned} \tag{23}$$

where we used the variational estimate (20). We define norms such that we can abbreviate the above result as

$$\|(f_0,\psi_0)\|_{\mathscr{B}^{4,3}} \le C\|\mathring{\Pi}(\Psi(g_0,E_0))\|_{\mathscr{B}^{0,1}}. \tag{24}$$

3.1.6. *Iteration*

If we let $(h_0, X_0) = \rho D\Psi^*_{(g_0,E_0)}(f_0,\psi_0)$ and define

$$\|(h,X)\|_{\mathscr{B}^{2,2}} = \|(h,X)\|_{C^{2,\alpha}_{\phi,\phi^{\frac{7}{2}}\rho^{-\frac{1}{2}}}(\Omega)} + \|(h,X)\|_{L^2_{\rho^{-1}}(\Omega)},$$

it follows from (24) that

$$\|(h_0,X_0)\|_{\mathscr{B}^{2,2}} \le C\|(f_0,\psi_0)\|_{\mathscr{B}^{4,3}} \le C'\|\mathring{\Pi}(\Psi(g_0,E_0))\|_{\mathscr{B}^{0,1}}.$$

The point is that by design, $\mathring{\Pi}(\Psi(g_0+h_0,E_0+X_0))$ is quadratic in (h_0,X_0), i.e. in $\mathring{\Pi}(\Psi(g_0,E_0))$. We set up a recursion by a quasi-Newtonian scheme to determine a sequence $(h_m,X_m) = \rho D\Psi^*_{(g_0,E_0)}(f_m,\psi_m)$, satisfying

$$\mathring{\Pi}(D\Psi_{(g_0,E_0)}\rho D\Psi^*_{(g_0,E_0)}(f_{m+1},\psi_{m+1})) = -\mathring{\Pi}(\Psi(g_m,E_m)),$$

where $g_m = g_0 + \sum_{k=0}^{m-1} h_k$, and $E_m = E_0 + \sum_{k=0}^{m-1} X_k$. Note that the linearization is computed at the starting data, which we take to be smooth.

For sufficiently small $\Psi(g_0,E_0)$, the resulting convergence is not quadratic, but it is superlinear; for more details, see [55, Section 6.3 and Appendix D], cf. [22,23], and [21, Appendix G]. We obtain $g_m \to g = g_0 + h$ and $E_m \to E = E_0 + X$, with $\mathring{\Pi}(\Psi(g,E)) = 0$, in other words, $\Psi(g,E) \in \zeta\mathscr{K}$. Here $(h,X) = \rho D\Psi^*_{(g_0,E_0)}(f,\psi) \in \mathscr{B}^{2,2}$, where $f = \sum_{k=0}^{\infty} f_k$ and $\psi = \sum_{k=0}^{\infty}\psi_k$, with $(f,\psi) \in \mathscr{B}^{4,3} \cap S_{g_0}$, and $\|(h,X)\|_{\mathscr{B}^{2,2}} \le C\|\mathring{\Pi}(\Psi(g_0,E_0))\|_{\mathscr{B}^{0,1}}$. By elliptic bootstrapping on the quaslinear equation $\Psi(g,E) \in \zeta\mathscr{K}$, one can show that (h,X) extends smoothly by zero across $\partial\Omega$.

3.1.7. *Handling the cokernel*

Thus far we have not glued anything, we have just referenced data near $(\mathring{g},0)$ on a domain Ω. We will now restrict to $\Omega = \mathscr{A}_1$, the unit annulus. Given asymptotically flat initial data (g,E) satisfying $\Psi(g,E) = 0$, we use a cutoff function to smoothly patch (g,E) to $(g_{\mathrm{RN}},E_{\mathrm{RN}})$ in $\mathscr{A}_R$, obtaining $(\hat{g}_R,\hat{E}_R)$. We may assume for exposition that $g_{ij}(x) - \delta_{ij} = O(|x|^{-1})$ and $E^i(x) = O(|x|^{-2})$ (and derivatives used in the proof attain additional decay), though we could formulate somewhat more generally. For instance, we could pose the starting solution of the constraints is asymptotically flat of order q, i.e. in asymptotic coordinates $|x| > 1$, $\partial_x^\alpha(g_{ij}(x) - \delta_{ij}) = O(|x|^{-q-|\alpha|})$ and $\partial_x^\beta E^i = O_1(|x|^{-1-q-|\beta|})$ for $q > 1/2$, $|\alpha| \le 2$ and $|\beta| \le 1$. Notice that $R(g) = 2\kappa|E|^2_g = O(|x|^{-2-2q})$ which is integrable, and hence the ADM mass exists. In any case, we consider $q = 1$ for exposition, and the interested reader can readily generalize.

The exterior we attach depends on the parameters m, Q, as well as $c \in \mathbb{R}^3$ about which we center $(g_{\mathrm{RN}},E_{\mathrm{RN}})$. We then pullback to the unit annulus, as follows: for $x \in \mathscr{A}_1$, let $(\tilde{g}_R)_{ij}(x) = (\hat{g}_R)_{ij}(Rx)$, and $\tilde{E}^i_R(x) = R\hat{E}^i_R(Rx)$. Observe that $(\tilde{g}_R,\tilde{E}_R)$ differs from $(\mathring{g},0)$ in $\mathscr{A}_1$ by $O(R^{-1})$, while $\Psi(\tilde{g}_R,\tilde{E}_R)(x) = R^2\Psi(\hat{g}_R,\hat{E}_R)(Rx) = O(R^{-1})$. We see the scaling was chosen so that we can solve the vacuum constraints on $\mathscr{A}_1$ and simply scale back. Moreover, for all large R, we can apply the previous analysis to find (h_R,X_R) that extend smoothly by zero outside $\mathscr{A}_1$, such that with $(\overline{g}_R,\overline{E}_R) = (\tilde{g}_R + h_R, \tilde{E}_R + X_R)$, we have $\mathring{\Pi}(\Psi(\overline{g}_R,\overline{E}_R)) = 0$.

The final step is to arrange $\Psi(\overline{g}_R,\overline{E}_R) = 0$, for which we recall $\mathscr{K} = \mathrm{span}\{1,x^1,x^2,x^3\} \oplus \mathrm{span}\{1\}$. Projecting $\Psi(\overline{g}_R,\overline{E}_R)$ onto the second summand yields (by scaling)

$$R\int_{\mathscr{A}_1} \mathrm{div}_{\overline{g}_R}\overline{E}_R\,\mathrm{d}\mu_{\overline{g}_R} = R\int_{\partial\mathscr{A}_1}\tilde{E}_R\cdot\nu_{\tilde{g}_R}\,\mathrm{d}\sigma_{\tilde{g}_R} = \int_{\{|x|=2R\}} E_{\mathrm{RN}}\cdot\nu_{g_{\mathrm{RN}}}\,\mathrm{d}\sigma_{g_{\mathrm{RN}}} - \int_{\{|x|=R\}} E\cdot\nu_g\,\mathrm{d}\sigma_g.$$

Since $\mathrm{div}_g E = 0$, the integral $\int_{\{|x|=R\}} E\cdot\nu_g\,\mathrm{d}\sigma_g$ is independent of $R > 1$, and the value is $4\pi Q$. Thus if we patch the solution to Reissner–Nordström data with charge Q, then we conclude $\int_{\mathscr{A}_1}\mathrm{div}_{\overline{g}_R}\overline{E}_R\,\mathrm{d}\mu_{\overline{g}_R} = 0$. Since $\mathrm{div}_{\overline{g}_R}\overline{E}_R = b\zeta$ for a constant b, we conclude $b = 0$, and hence $\mathrm{div}_{\overline{g}_R}\overline{E}_R = 0$. The reader might compare this with our earlier remark in Section 1.2.1 about how to prescribe the electric field data.

For the rest of the section, we extend the summation convention to all pairs of repeated indices (even if both are down). Since in Cartesian coordinates, $DR_{\mathring{g}}(h) = h_{ij,ij} - h_{ii,jj}$, we see that by simple expansion of the scalar curvature term

$$\begin{aligned}\int_{\mathscr{A}_1} (R(\overline{g}_R) - 2\kappa|\overline{E}_R|^2_{\overline{g}_R})\,\mathrm{d}x &= \int_{\partial\mathscr{A}_1} ((\tilde{g}_R)_{ij,i} - (\tilde{g}_R)_{ii,j})\,\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma} + O(R^{-2})\\ &= R^{-1}\int_{\partial\mathscr{A}_R} ((\hat{g}_R)_{ij,i} - (\hat{g}_R)_{ii,j})\,\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma} + O(R^{-2})\\ &= R^{-1}16\pi(m - m(g)) + O(R^{-2}).\end{aligned}$$

So the projection is to leading order controlled by the mass m of the Reissner–Nordström exterior.

The projection onto the linear elements requires a bit more attention. For the metric g, we find that for any $r > r_0 > 1$,

$$\int_{\{r_0\le|x|\le r\}} x^k(g_{ij,ij} - g_{ii,jj})\,\mathrm{d}x = \left(\int_{\{|x|=r\}} - \int_{\{|x|=r_0\}}\right)(x^k(g_{ij,i} - g_{ii,j}) - (\delta^k_i g_{ij} - \delta^k_j g_{ii}))\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma}.$$

Note that in the last terms of the integral on the right, we could replace the terms of the form $g_{\ell m}$ with $g_{\ell m} - \delta_{\ell m}$. Moreover, *using the constraint equations*, the integrand $(g_{ij,ij} - g_{ii,jj}) = O(|x|^{-4})$, and so the left hand side is $O(\log r)$.

If instead the above is evaluated at the Reissner–Nordström data, by using the asymptotic parity and the fact that the center of mass does not appear in the $|x|^{-1}$ term in the expansion, we can show that for $|c| \le r/2$ and for m bounded, $\int_{\{|x|\ge r\}} x^k((g_{\mathrm{RN}})_{ij,ij} - (g_{\mathrm{RN}})_{ii,jj})\,\mathrm{d}x = O(r^{-1})$, on the one hand, while on the other

$$\lim_{r\to\infty}\int_{\{|x|=r\}} (x^k((g_{\mathrm{RN}})_{ij,i} - (g_{\mathrm{RN}})_{ii,j}) - (\delta^k_i(g_{\mathrm{RN}})_{ij} - \delta^k_j(g_{\mathrm{RN}})_{ii}))\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma} = 16\pi m c^k.$$

Assembling all of this, noting that the rescaled exterior is a Reissner–Nordström with center $\tilde{c} = cR^{-1}$, we obtain

$$\begin{aligned}&\int_{\mathscr{A}_1} x^k(R(\overline{g}_R) - 2\kappa|\overline{E}_R|^2_{\overline{g}_R})\,\mathrm{d}x\\ &= \int_{\partial\mathscr{A}_1} x^k((\tilde{g}_R)_{ij,i} - (\tilde{g}_R)_{ii,j})\,\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma} + O(R^{-2})\\ &= R^{-2}\int_{\{|x|=2R\}} (x^k((g_{\mathrm{RN}})_{ij,i} - (g_{\mathrm{RN}})_{ii,j}) - (\delta^k_i(g_{\mathrm{RN}})_{ij} - \delta^k_j(g_{\mathrm{RN}})_{ii}))\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma}\\ &\quad - R^{-2}\int_{\{|x|=R\}} (x^k(g_{ij,i} - g_{ii,j}) - (\delta^k_i g_{ij} - \delta^k_j g_{ii}))\mathring{\nu}^j\,\mathrm{d}\mathring{\sigma} + O(R^{-2})\\ &= R^{-1}\left(16\pi m\tilde{c}^k + O\left(\frac{\log R}{R}\right)\right) + O(R^{-2}).\end{aligned} \tag{25}$$

Thus we get a map

$$(m,\tilde{c}) \mapsto \frac{R}{16\pi}\int_{\mathscr{A}_1} (1,x)(R(\overline{g}_R) - 2\kappa|\overline{E}_R|^2_{\overline{g}_R})\,\mathrm{d}x = (m - m(g) + \xi(m,\tilde{c}), m\tilde{c} + \Xi(m,\tilde{c})),$$

where $\xi(m,\tilde{c}) = O(R^{-1})$ and $\Xi(m,\tilde{c}) = O(\log R/R)$. Moreover, the construction is *continuous* with respect to m and $\tilde{c}$, and so ξ and Ξ are continuous; this is fairly straightforward but cumbersome to prove (see, e.g. [17, Section A.7]). A fixed-point or degree argument can finish the proof from here. For example, the function

$$F(m,\tilde{c}) = \left(m(g) - \xi(m,\tilde{c}), -\frac{\Xi(m,\tilde{c})}{m(g) - \xi(m,\tilde{c})}\right),$$

is readily seen to map $\{(m,\tilde{c}) : |m - m(g)| \le |m(g)|/2, |\tilde{c}| \le 1/2\}$ to itself, and the Brouwer fixed point theorem can then be applied. The fixed point gives parameters that solve the problem.

We remark that we did not assume the original metric g enjoyed enough parity symmetry to admit a well-defined center of mass. If this were the case, we could argue that the resulting center $c = R\tilde{c}$ is close to that of the original, cf. [22, 23].

3.2. *Further applications of deformation*

Suppose the starting data, say (g_0, E_0), is such that $D\Psi^*_{(g_0,E_0)}$ admits only trivial kernel in Ω. The same perturbative analysis above can be applied, the only difference is that since there is no kernel to deal with, we do not need the projected operator, and so the technique yields a localized deformation result: given a compactly contained subdomain of Ω and sufficiently small compactly supported S and σ, say, there are h and X which smoothly extend by zero across $\partial\Omega$ with $\Psi(g_0 + h, E_0 + X) = \Psi(g_0, E_0) + (S, \sigma)$. Similar comments apply to scalar curvature deformation [22], as well as for the general constraints case [23], though in the latter case maintaining the dominant energy condition at the borderline is somewhat subtle [55].

As mentioned earlier, one can combine conformal gluing and localized gluing to obtain connected sum constructions leaving the original data intact outside the gluing neck [21, 24, 25]. The additional assumption needed, beyond any needed for the conformal gluing, is that the corresponding operator for the localized gluing problem has trivial kernel. If you have a sequence of solutions of the nonlinear problem (vacuum constraints, or a scalar curvature equation, say) which converges to a solution for which the linearized adjoint has trivial kernel in Ω, then you can *glue back in* the original data inside some appropriate subdomain, and use the localized perturbations (unobstructed in this case) to reimpose the nonlinear equation.

For an interesting application combining localized deformations of constraint data with conformal deformations, consider an asymptotically flat metric g with nonnegative scalar curvature $R(g)$, which is not *static* outside of some bounded region $\Omega \subset \mathbb{R}^3$ (in the sense that the adjoint DR^*_g of the linearization of the scalar curvature has trivial kernel in the exterior $\mathbb{R}^3 \setminus \overline{\Omega}$). It is possible using localized deformations to find a compactly contained domain $V \subset R^3 \setminus \overline{\Omega}$ such that for S compactly supported in V, and for $\varepsilon > 0$ sufficiently small, there is a metric $\bar{g}_\varepsilon$ with $R(\bar{g}_\varepsilon) = R(g) + \varepsilon S$, and $\bar{g}_\varepsilon - g = O(\epsilon)$ is supported in V. If $S \geq 0$ and is nontrivial, one can then use an elementary conformal transformation to *bring the ADM mass down* while keeping the scalar curvature nonnegative, with conformal factor $u_\varepsilon > 0$ which is one in a neighborhood of $\overline{\Omega}$ [57]. In terms of asymptotically flat extensions of a given region, then, this shows that unless the data outside the region were static in the above sense, there would be extensions with nonnegative scalar curvature and lower ADM mass, and hence in particular the original configuration does not minimize the Bartnik mass. To summarize, a localized deformation is used to bump the scalar curvature up (adding energy density) in a compact set outside the region $\overline{\Omega}$, which does not change the ADM mass, and then a conformal deformation removes some of this scalar curvature and in so doing brings down the ADM mass, leaving the original geometry of $\overline{\Omega}$ intact.

Similar considerations hold for the general (non-time-symmetric) case, though maintaining the dominant energy condition seems to require a modification of the constraints operator to have trivial kernel [55, 58]. An interpretation of the kernel elements for the modified operator, akin to that in [51], and has been established in the recent work [59] on constraint deformations and the dominant energy condition, with striking applications to the Bartnik mass.

Another application is the gluing of initial data to *interpolate* their scalar curvatures [60], or more generally interpolating (μ, J) from the constraints operator [55]. As an interesting example, one can glue initial data for a Kerr solution (including Schwarzschild) to a different Kerr solution across an annular region where the dominant energy condition can be maintained (see Remark after Theorem 1.4 in [55]).

Finally, we discussed earlier the use of the conformal method to achieve solutions with certain prescribed asymptotic behavior, with stability for the Einstein evolution in mind. Motivated by questions about the asymptotics posed in [16], the work [61] proceeds to seed the desired asymptotics, and then correct to solutions of the constraints, but in a different manner than that of [47, 48], in part employing variational methods of the kind used in [16, 23].

3.3. *Alternate approach to solving the projected problem*

We review the steps in the procedure shown above for gluing to an asymptotically flat model family. We start by patching initial data to a member of the model family, with the following parameters at our disposal: R (measuring the scale), the exterior mass m and center of mass c; for the Einstein–Maxwell constraints, we match the electric charge Q. For large enough R the patched data is sufficiently close to flat, and we solve the constraints up to a finite-dimensional obstruction space. To accomplish this, we iterate linear corrections, in spaces with suitable decay at the boundary of the annulus. The procedure is controlled using basic elliptic estimates, both integral and pointwise, and with suitable weights. The integral estimates for the time-symmetric case are relatively straightforward, while the weighted Schauder estimates follow from the standard interior estimates and scaling. From here, a topological argument using the finite-dimensional space of parameters allows us to cover the cokernel and solve the constraints. Similar comments apply in the non-time-symmetric case.

The main first step, then, is how to solve the relevant linear equations with compact support. For $g = \mathring{g} + h$ with h small, the scalar curvature to leading order is governed by a divergence:

$$R(\mathring{g} + h) = DR_{\mathring{g}}(h) + \mathcal{Q}(h) = \mathrm{div}_{\mathring{g}}\mathrm{div}_{\mathring{g}}(h - (\mathrm{tr}_{\mathring{g}} h)\mathring{g}) + \mathcal{Q}(h),$$

where $\mathcal{Q}(h)$ is a combination with bounded coefficients of terms of components of $h \otimes \partial^2 h$ and $\partial h \otimes \partial h$. A similar expansion holds for the constraints operator. At the linear level, then, solving for a compactly supported perturbation for the constraints map near $(\mathring{g}, 0)$ amounts to being able to solve suitable divergence equations with compact support.

Recent works [62, 63] have indeed developed this approach near the flat data (with an extension to data near hyperbolic data obtained in [64]). For instance, to solve a double divergence equation, the authors use a Bogovskiĭ-type operator as follows. Let $\Omega \subset \mathbb{R}^3$ be an open set, which for simplicity we take to be convex, and containing an open ball B, and let $\eta \in C_c^\infty(B)$ with $\int_{\mathbb{R}^3} \eta(x)\,\mathrm{d}x = 1$. Then for any $f \in C_c^\infty(\mathbb{R}^n)$ we define the tensor Sf to have Cartesian components

$$(Sf)^{ij}(x) = \int_{\mathbb{R}^3} \left(\int_{|x-y|}^{\infty} \eta\left(r\frac{x-y}{|x-y|} + y \right) r^2\,\mathrm{d}r \right) \frac{(x-y)^i (x-y)^j}{|x-y|^3} f(y)\,\mathrm{d}y.$$

If $f \in C_c^\infty(\Omega)$, then Sf is supported in Ω as well, and moreover, if $\int_\Omega f(x)\,\mathrm{d}x = 0 = \int_\Omega x^k f(x)\,\mathrm{d}x$, for $k \in \{1,2,3\}$, then $\mathrm{div}_{\mathring{g}}\mathrm{div}_{\mathring{g}}(Sf) = f$. The integral conditions are of course readily seen to be the orthogonality of f to the cokernel of the double divergence. One can furthermore use a straightforward finite patching argument to handle more general domains such as the annuli used in gluing constructions.

As a remark related to this and to the Einstein–Maxwell equations, we recall the well-known procedure for solving the divergence equation $\mathrm{div}_{\mathring{g}} X = f$, with f and the solution X compactly supported (see, e.g. [65, Section 1.3]). Of course we assume that $\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = 0$. In one dimension this is straightforward, as $X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ satisfies the required conditions. We can proceed inductively. Suppose f is compactly supported in $D = D_n \times I \subset \mathbb{R}^{n+1}$, where D_n is an open rectangle in $\mathbb{R}^n$ and $I \subset \mathbb{R}$ is an open interval. For $x \in \mathbb{R}^n$, we let $F(x) = \int_{\mathbb{R}} f(x,t)\,\mathrm{d}t$. By assumption, then $\int_{\mathbb{R}^n} F(x)\,\mathrm{d}x = 0$, so by induction, there are $Y^1, \ldots, Y^n$ supported in D_n with

$\partial_i Y^i(x) = F(x)$. If θ is supported in I with $\int_{\mathbb{R}} \theta(t)\,\mathrm{d}t = 1$, we let $X^i(x, x^{n+1}) = Y^i(x)\theta(x^{n+1})$ for $i \in \{1, \dots, n\}$ and

$$X^{n+1}(x, x^{n+1}) = \int_{-\infty}^{x^{n+1}} (f(x,t) - \theta(t)F(x))\,\mathrm{d}t.$$

The vector field X with components $X^1, \dots, X^{n+1}$ is supported in D and has divergence f.

Returning to the constraints, one would still need to solve the nonlinear projected problem, which is essentially done by an iteration scheme. That said, one major benefit of the Bogovskiĭ-type operator is that regularity is maintained when the data is in appropriate Sobolev spaces, a question which arises when considering the natural regularity for initial data for the evolution problem. Moreover the method has been extended beyond the asymptotically flat regime to give an asymptotically hyperboloidal construction [64], and this is certainly intriguing since the asymptotically hyperboloidal case proves to be more challenging (for instance the conformal infinity is not a point as it is for the asymptotically Euclidean case), see for example [66–68]. The Bogovskiĭ-type operator method has also been used for other localization results such as Carlotto–Schoen type gravitational shielding constructions [62]. Finally, one often wants to connect the solutions of the Einstein constraints to the spacetimes they determine, whether in theory or numerically. Numerical implementation of gluing constructions seems quite challenging, and though there have been a handful of works in this direction [69–72], alternative methods to do gluing constructions could prove useful if they are more readily implemented numerically. It is natural to wonder if the Bogovskiĭ-type operator approach of [63] might be helpful in this direction, and likewise for the characteristic gluing constructions of [11] involving transport-elliptic differential equations, which are of a different character than the spacelike gluing.

4. Further applications to spacetimes

As mentioned earlier, one of the early goals for the gluing construction was to create vacuum initial data which would evolve to a vacuum spacetime that is asymptotically simply in the sense of Penrose conformal compactification, a *purely radiative* spacetime. Since work of Friedrich [32] the following was known: if one could construct a family of solutions to the vacuum constraints with small mass tending to zero, each of which is Schwarzschild in a *uniform* neighborhood of spacelike infinity, then for members of the family with sufficiently small mass, the spacetime evolution would be an asymptotically simple vacuum spacetime. The result in [22] on its own did not quite give such a construction. The main issue, as one can see from (25), is that as the mass m tends to zero, the projection onto the linear directions is on the order of mc, and so for small mass, the center of mass may need to grow larger to accommodate. To control the center, Chruściel and Delay [30] restrict to a family of parity-symmetric metrics, for which the center of mass in the construction can always be taken to vanish. In [31], we construct a larger family of solutions to the constraints with small masses and controlled centers, by using the second variation of the mass at the Euclidean data [73, 74] to control the mass from below in terms of the center. With either approach, the gluing to Schwarzschild at infinity then finishes the initial data construction.

In a completely different direction, Li and Yu [75] construct vacuum initial data, free of trapped surfaces, whose evolution will contain a trapped surface (cf. [76]). Of note, the solution contains an interior region of Minkowskian data, while the exterior is an exterior of a space-time slice of a Kerr spacetime. The construction builds on the monumental work of Christodoulou on formation of trapped surfaces *in vacuum* [77], from data on null hypersurfaces; in particular, the construction in [75] uses data given at past null infinity.

While we do not have the space and time to discuss any details here, there is one feature of the proof that we want to note. The construction uses gluing to Kerr exterior as in [23], but in a regime where the limiting data is not Minkowksian, but Kerrian. As such, the cokernel has *smaller* dimension than at the flat data, so the gluing is not quite as obstructed. This has been observed in other places, such as in initial data constructions with positive cosmological constant [78] and in more recent work on *multi-localized* initial data [17]. In the latter work, the relevant approximate kernel is that of a Schwarzschild, which is one-dimensional (for non-zero mass). The motivation for the multi-localized initial data constructed in [17] comes from the nonlinear stability result of Anderson and Pasqualotto [79]. The template for the construction is the Brill–Lindquist metric (5), and we glue ends on around each of the punctures in the set $\mathfrak{p}$ to make initial data on $\mathbb{R}^3$. Of note is that we are able to handle certain configurations where $\mathfrak{p}$ is infinite, and in particular, we construct smooth metrics on $\mathbb{R}^3$ with zero scalar curvature and infinitely many minimal spheres, each of them exterior to all the others.

As remarked earlier, in recent years *characteristic gluing* has been developed in a series of papers by Aretakis, Czimek and Rodnianski; for an overview and further references, see [11]. For the characteristic initial value problem, one specifies certain free data on what will be two transversely intersecting null hypersurfaces in the vacuum spacetime, after which one can appeal to local well-posedness results for the characteristic initial value problem for the Einstein vacuum equation [80, 81]. For the characteristic gluing problem, one seeks to glue data on two spheres along a null hypersurface. As with the spacelike case, the relevant equations that must be satisfied arise from the vacuum Einstein equation together with the fundamental compatibility equations for an embedding, in this case the compatibility coming from a double null foliation. The resulting equations for metric and curvature components are the *null structure equations*, of transport-elliptic character along either the ingoing or outgoing null hypersurface. As is noted in [11], there are obstructions to gluing data from two spheres, such as a monotonicity deriving from the Raychaudhuri equation.

There are a number of noteworthy applications of their technique, and we just mention a few. Their methods can recover the asymptotic gluing to Kerr as in [23], as well as localized gluing from [16], and without loss of decay in the transition region. In [82], the method is employed to achieve *obstruction-free* gluing. The asymptotic gluing construction along the lines of [22, 23] glues initial data with certain energy–momenta to Kerr data with energy–momenta which is close to the starting one, as we say above, tuned to cover the linear obstruction from the approximate cokernel. A natural question is whether one can glue pieces from spacetimes whose parameters are not necessarily close, and results in this direction have been obtained in [82].

The characteristic gluing construction has been employed by Kehle and Unger [83] to perform a specific *event horizon gluing* in the context of the Einstein–Maxwell charged scalar field system, in spherical symmetry. As an upshot, they produce a counterexample to the third law of black hole thermodynamics.

5. Conclusion

Gluing methods for constructing solutions of the Einstein constraint equations have garnered a number of noteworthy achievements. While more than twenty years have passed since the gluing results of [22] and [20], new techniques and applications are still being developed and continue to highlight the interplay between initial data results and the evolution problem.

Dedication

This article is dedicated to Mme. Yvonne Choquet-Bruhat, for giving so much to so many of us for so many years. The memory of Yvonne that is emblazoned upon me comes from over twenty

years ago at a workshop in Cargèse, Corsica, marking the fiftieth anniversary of her landmark paper on the initial-value problem. Walking back to the seminar room with a colleague, we encountered Yvonne taking a bit of respite from the sunshine under a tree; while it was not an apple tree, what sprung to my imagination was the universe (or the spirit of Sir Isaac Newton) willing one to fall and land in the palm of her hand.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Acknowledgments

The author would like to thank Farhan Abedin, David A. Maxwell and Iva Stavrov Allen for their comments on a draft manuscript. The author is supported in part by the AMS-Simons Research Enhancement Grant for PUI Faculty.

References

[1] Y. Choquet-Bruhat, *General Relativity and the Einstein Equations*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2009, xxvi + 785 pages.

[2] Y. Choquet-Bruhat, "The Cauchy problem", in *Gravitation: An Introduction to Current Research* (L. Witten, ed.), John Wiley & Sons, Inc., New York-London, 1962, p. 130-168.

[3] Y. Choquet-Bruhat, J. W. York Jr, "The Cauchy problem", in *General Relativity and Gravitation* (A. Held, ed.), Plenum Press, New York-London, 1980, p. 99-172.

[4] Y. Choquet-Bruhat, "Beginnings of the Cauchy problem for Einstein's field equations", *Surv. Differ. Geom.* **20** (2015), p. 1-16.

[5] Y. Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), p. 141-225.

[6] J. W. York Jr, "Gravitational degrees of freedom and the initial-value problem", *Phys. Rev. Lett.* **26** (1971), p. 1656-1658.

[7] D. Maxwell, "Initial data in general relativity described by expansion, conformal deformation and drift", *Comm. Anal. Geom.* **29** (2021), no. 1, p. 207-281.

[8] A. Lichnerowicz, "L'intégration des équations de la gravitation relativiste et le problème des n corps", *J. Math. Pures Appl.* **23** (1944), no. 9, p. 37-63.

[9] P. Hintz, "Gluing small black holes along timelike geodesics I: formal solution", preprint, 2023, https://arxiv.org/abs/2306.07409.

[10] P. Hintz, "Gluing small black holes into initial data sets", *Commun. Math. Phys.* **405** (2024), no. 5, article no. 114.

[11] S. Aretakis, S. Czimek, I. Rodnianski, "The characteristic gluing problem for the Einstein equations and applications", preprint, 2022, https://arxiv.org/abs/2107.02441.

[12] P. T. Chruściel, J. Corvino, J. Isenberg, "Initial data for the relativistic gravitational N-body problem", *Class. Quantum Gravity* **27** (2010), no. 22, p. 6.

[13] P. T. Chruściel, R. Mazzeo, "On "many-black-hole" vacuum spacetimes", *Class. Quantum Gravity* **20** (2003), no. 4, p. 729-754.

[14] P. T. Chruściel, J. Corvino, J. Isenberg, "Construction of N-body time-symmetric initial data sets in general relativity", in *Complex Analysis and Dynamical Systems IV. Part 2*, Contemporary Mathematics, vol. 554, American Mathematical Society, Providence, RI, 2011, p. 83-92.

[15] P. T. Chruściel, J. Corvino, J. Isenberg, "Construction of N-body initial data sets in general relativity", *Commun. Math. Phys.* **304** (2011), no. 3, p. 637-647.

[16] A. Carlotto, R. Schoen, "Localizing solutions of the Einstein constraint equations", *Invent. Math.* **205** (2016), no. 3, p. 559-615.

[17] J. Anderson, J. Corvino, F. Pasqualotto, "Multi-localized time-symmetric initial data for the Einstein vacuum equations", *J. Reine Angew. Math. (Crelle's Journal)* **808** (2024), p. 67-110.

[18] C. Racine, *Le problème des N corps dans la théorie de la relativité*, Thèse, Gauthier-Villars, Paris, 1934, http://www.numdam.org/item/THESE_1934__158__1_0.pdf.

[19] G. Darmois, "Les équations de la gravitation einsteinienne", *Mémor. Sci. Math.* **25** (1927), article no. 58.
[20] J. Isenberg, R. Mazzeo, D. Pollack, "Gluing and wormholes for the Einstein constraint equations", *Commun. Math. Phys.* **231** (2002), no. 3, p. 529-568.
[21] P. T. Chruściel, E. Delay, "On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications", *Mém. Soc. Math. Fr. (N.S.)* **94** (2003), p. vi + 103.
[22] J. Corvino, "Scalar curvature deformation and a gluing construction for the Einstein constraint equations", *Commun. Math. Phys.* **214** (2000), no. 1, p. 137-189.
[23] J. Corvino, R. M. Schoen, "On the asymptotics for the vacuum Einstein constraint equations", *J. Differential Geom.* **73** (2006), no. 2, p. 185-217.
[24] P. T. Chruściel, J. Isenberg, D. Pollack, "Initial data engineering", *Commun. Math. Phys.* **257** (2005), no. 1, p. 29-42.
[25] J. Corvino, M. Eichmair, P. Miao, "Deformation of scalar curvature and volume", *Math. Ann.* **357** (2013), no. 2, p. 551-584.
[26] L. Mazzieri, "Generalized gluing for the Einstein constraint equations", *Calc. Var. Partial Differ. Equ.* **34** (2009), no. 4, p. 453-473.
[27] E. Delay, L. Mazzieri, "Refined gluing for the vacuum Einstein constraint equations", *Geom. Dedicata* **173** (2014), p. 393-415.
[28] J. Isenberg, R. Mazzeo, D. Pollack, "On the topology of vacuum spacetimes", *Ann. Henri Poincaré* **4** (2003), no. 2, p. 369-383.
[29] I. Stavrov Allen, "A gluing construction regarding point particles in general relativity", *Ann. Henri Poincaré* **10** (2010), no. 8, p. 1437-1486.
[30] P. T. Chruściel, E. Delay, "Existence of non-trivial, vacuum, asymptotically simple spacetimes", *Class. Quantum Gravity* **19** (2002), no. 9, p. L71-L79.
[31] J. Corvino, "On the existence and stability of the Penrose compactification", *Ann. Henri Poincaré* **8** (2007), no. 3, p. 597-620.
[32] H. Friedrich, "On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure", *Commun. Math. Phys.* **107** (1986), p. 587-609.
[33] R. Bartnik, "Quasi-spherical metrics and prescribed scalar curvature", *J. Differential Geom.* **37** (1993), no. 1, p. 31-71.
[34] J. Isenberg, D. Maxwell, "A phase space approach to the conformal construction of non-vacuum initial data sets in general relativity", preprint, 2021, https://arxiv.org/abs/2106.15027v2.
[35] D. R. Brill, R. W. Lindquist, "Interaction energy in geometrostatics", *Phys. Rev. (2)* **131** (1963), p. 471-476.
[36] C. Sormani, I. Stavrov Allen, "Geometrostatic manifolds of small ADM mass", *Comm. Pure Appl. Math.* **72** (2019), no. 6, p. 1243-1287.
[37] C. W. Misner, "The method of images in geometrostatics", *Ann. Phys.* **24** (1963), p. 102-117.
[38] Y. Fourès-Bruhat, "Sur l'intégration des équations de la relativité génŕale", *J. Ration. Mech. Anal.* **5** (1956), no. 6, p. 951-966.
[39] N. Ó. Murchadha, J. W. York Jr, "Existence and uniqueness of solutions of the Hamiltonian constraint of general relativity on compact manifolds", *J. Math. Phys.* **14** (1973), no. 11, p. 1151-1557.
[40] N. Ó. Murchadha, J. W. York Jr, "Initial-value problem of general relativity. I. General formulation and physical formulation", *Phys. Rev. D* **10** (1974), no. 2, p. 428-436.
[41] D. Maxwell, "The conformal method and the conformal thin-sandwich method are the same", *Class. Quantum Gravity* **31** (2014), no. 14, article no. 145006.
[42] R. M. Schoen, S.-T. Yau, "The energy and the linear momentum of space-times in general relativity", *Commun. Math. Phys.* **79** (1981), no. 1, p. 47-51.
[43] R. Schoen, S. T. Yau, "On the proof of the positive mass conjecture in general relativity", *Commun. Math. Phys.* **65** (1979), no. 1, p. 45-76.
[44] H. L. Bray, "Proof of the Riemannian Penrose inequality using the positive mass theorem", *J. Differential Geom.* **59** (2001), p. 177-267.
[45] M. Eichmair, L.-H. Huang, D. A. Lee, R. M. Schoen, "The spacetime positive mass theorem in dimensions less than eight", *J. Eur. Math. Soc.* **18** (2016), p. 83-121.
[46] L.-H. Huang, "On the center of mass of isolated systems with general asymptotics", *Class. Quantum Gravity* **26** (2009), no. 1, article no. 015012.
[47] A. J. Fang, J. Szeftel, A. Touati, "Initial data for Minkowski stability with arbitrary decay", preprint, 2024, https://arxiv.org/abs/2401.14353.
[48] A. J. Fang, J. Szeftel, A. Touati, "Spacelike initial data for black hole stability", preprint, 2024, https://arxiv.org/abs/2405.02071.
[49] A. E. Fischer, J. E. Marsden, "Deformations of the scalar curvature", *Duke Math. J.* **42** (1975), no. 3, p. 519-547.
[50] A. E. Fischer, J. E. Marsden, "Linearization stability of the Einstein equations", *Bull. Am. Meteorol. Soc.* **79** (1973), p. 997-1003.

[51] V. Moncrief, “Spacetime symmetries and linearization of the Einstein equations I”, *J. Math. Phys.* **16** (1975), p. 493-498.

[52] J. Isenberg, D. Maxwell, D. Pollack, “A gluing construction for non-vacuum solutions of the Einstein constraint equations”, *Adv. Theor. Math. Phys.* **9** (2005), no. 1, p. 129-172.

[53] L. Bieri, P. T. Chruściel, “Future-complete null hypersurfaces, interior gluing and the Trautman-Bondi mass”, in *Nonlinear Analysis in Geometry and Applied Mathematics* (L. Bieri, P. T. Chruściel, S.-T. Yau, eds.), Harvard University Center of Mathematical Sciences and Applications Series in Mathematics, vol. 1, International Press, Somerville, MA, 2017, p. 1-31.

[54] P. T. Chruściel, “Long time existence from interior gluing”, *Class. Quantum Gravity* **34** (2017), no. 14, article no. 145016.

[55] J. Corvino, L.-H. Huang, “Localized deformation for initial data sets with the dominant energy condition”, *Calc. Var. Partial Differ. Equ.* **59** (2020), no. 1, article no. 42.

[56] A. Douglis, L. Nirenberg, “Interior estimates for elliptic systems of partial differential equations”, *Comm. Pure Appl. Math.* **8** (1955), p. 503-538.

[57] J. Corvino, “A note on the Bartnik mass”, in *Nonlinear Analysis in Geometry and Applied Mathematics* (L. Bieri, P. T. Chruściel, S.-T. Yau, eds.), Harvard University Center of Mathematical Sciences and Applications Series in Mathematics, vol. 1, International Press, Somerville, MA, 2017, p. 49-75.

[58] J. Corvino, “A short note on the Bartnik mass”, *Pure Appl. Math. Q.* **15** (2019), no. 3, p. 827-838.

[59] L.-H. Huang, D. A. Lee, “Bartnik mass minimizing initial data sets and improvability of the dominant energy scalar”, *J. Differential Geom.* **126** (2024), no. 2, p. 741-800.

[60] E. Delay, “Localized gluing of Riemannian metrics in interpolating their scalar curvature”, *Differ. Geom. Appl.* **29** (2011), no. 3, p. 433-439.

[61] P. G. LeFloch, T.-C. Nguyen, “The seed-to-solution method for the Einstein constraints and the asymptotic localization problem”, *J. Funct. Anal.* **285** (2023), no. 9, article no. 110106.

[62] Y. Mao, Z. Tao, “Localized initial data for Einstein equations”, preprint, 2022, https://arxiv.org/abs/2210.09437.

[63] Y. Mao, S.-J. Oh, Z. Tao, “Initial data gluing in the asymptotically flat regime via solution operators with prescribed support properties”, preprint, 2023, https://arxiv.org/abs/2308.13031.

[64] P. T. Chruściel, A. Cogo, A. Nützi, “A Bogovskiĭ-type operator for Corvino–Schoen hyperbolic gluing”, preprint, 2024, https://arxiv.org/abs/2409.07502.

[65] L. Nirenberg, in *Topics in Nonlinear Functional Analysis (Chapter 6 by E. Zehnder. Notes by R. A. Artino. Revised reprint of the 1974 original)*, Courant Lecture Notes in Mathematics, vol. 6, New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2001, p. xii + 145.

[66] P. T. Chruściel, E. Delay, “Gluing constructions for asymptotically hyperbolic manifolds with constant scalar curvature”, *Comm. Anal. Geom.* **17** (2009), no. 2, p. 343-381.

[67] J. Isenberg, J. M. Lee, I. Stavrov Allen, “Asymptotic gluing of asymptotically hyperbolic solutions to the Einstein constraint equations”, *Ann. Henri Poincaré* **11** (2010), no. 5, p. 881-927.

[68] P. T. Allen, J. Isenberg, J. M. Lee, I. Stavrov Allen, “Asymptotic gluing of shear-free hyperboloidal initial data sets”, *Ann. Henri Poincaré* **22** (2021), no. 3, p. 771-819.

[69] D. Giulini, G. Holzegel, “Corvino’s construction using Brill waves”, preprint, 2005, https://arxiv.org/abs/gr-qc/0508070.

[70] G. Doulis, O. Rinne, “Numerical construction of initial data for Einstein’s equations with static extensions to spacelike infinity”, *Class. Quantum Gravity* **33** (2016), no. 7, article no. 075014.

[71] B. Daszuta, J. Frauendiener, “Numerical initial data deformation exploiting a gluing construction: I. Exterior asymptotic Schwarzschild”, *Class. Quantum Gravity* **36** (2019), no. 18, article no. 185008.

[72] D. Pook-Kolb, D. Giulini, “Numerical approach for Corvino-type gluing of Brill–Lindquist initial data”, *Class. Quantum Gravity* **36** (2019), no. 4, article no. 045011.

[73] Y. Choquet-Bruhat, J. E. Marsden, “Solution of the local mass problem in general relativity”, *Commun. Math. Phys.* **51** (1976), no. 3, p. 283-296.

[74] D. R. Brill, S. Deser, “Variational methods and positive energy in general relativity”, *Ann. Phys.* **50** (1968), p. 548-570.

[75] J. Li, P. Yu, “Construction of Cauchy data of vacuum Einstein field equations evolving to black holes”, *Ann. Math. (2)* **181** (2015), no. 2, p. 699-768.

[76] J. Li, H. Mei, “A construction of collapsing spacetimes in vacuum”, *Commun. Math. Phys.* **378** (2020), no. 2, p. 1343-1389.

[77] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2009, x + 589 pages.

[78] P. T. Chruściel, D. Pollack, “Singular Yamabe metrics and initial data with exactly Kottler–Schwarzschild–de Sitter ends”, *Ann. Henri Poincaré* **9** (2008), no. 4, p. 639-654.

[79] J. Anderson, F. Pasqualotto, “Global stability for nonlinear wave equations with multi-localized initial data”, *Ann. PDE* **8** (2022), no. 2, article no. 19.

[80] A. D. Rendall, “Reduction of the characteristic initial value problem to the Cauchy problem and its application to the Einstein equations”, *Proc. R. Soc. Lond. A* **427** (1990), no. 1872, p. 221-239.

[81] J. Luk, “On the local existence for the characteristic initial value problem in general relativity”, *Int. Math. Res. Not. IMRN* **2012** (2012), no. 20, p. 4625-4678.

[82] S. Czimek, I. Rodnianski, “Obstruction-free gluing for the Einstein equations”, preprint, 2022, https://arxiv.org/abs/2210.09663.

[83] C. Kehle, R. Unger, “Gravitational collapse to extremal black holes and the third law of black hole thermodynamics”, preprint, 2022, https://arxiv.org/abs/2211.15742.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 519-542
https://doi.org/10.5802/crmeca.287

Review article / *Article de synthèse*

The stability of Minkowski space and its influence on the mathematical analysis of General Relativity

La stabilité de l'espace de Minkowski et son influence sur l'analyse mathématique de la Relativité Générale

Jacques Smulevici [a]

[a] Laboratoire Jacques-Louis Lions (LJLL), Sorbonne Université, CNRS and Institut Universitaire de France, France

E-mail: jacques.smulevici@sorbonne-universite.fr

Abstract. The study of the stability of Minkowksi space has had a tremendous influence on the mathematical analysis of General Relativity and in a wider context, on geometric hyperbolic partial differential equations. This article proposes an overview of the context, ideas and part of the legacy surrounding this subject, focusing particularly on the original proof of Christodoulou–Klainerman.

Résumé. L'étude de la stabilité de l'espace de Minkowski a eu une influence majeure sur l'analyse mathématique de la Relativité Générale et, plus largement, sur les équations aux dérivées partielles hyperboliques géométriques. Cet article propose un aperçu du contexte, des idées et d'une partie de l'héritage entourant ce sujet, en mettant particulièrement l'accent sur la démonstration originale de Christodoulou–Klainerman.

Keywords. Einstein equations, General Relativity, Asymptotic stability.

Mots-clés. Équations d'Einstein, Relativité Générale, Stabilité asymptotique.

Manuscript received 30 October 2024, revised 31 January 2025, accepted 6 February 2025.

1. Introduction

Since the work of Minkowski and its unification of space and time [1], Special Relativity is beautifully encoded in the geometry of the simplest Lorentzian manifold, the Minkowski space $(\mathbb{R}^{1+3}, \eta)$, where η is the Lorentzian metric whose matrix of components in the standard Cartesian coordinates (t, x, y, z) of $\mathbb{R}^{1+3}$ is given by $\left(\eta_{\alpha\beta}\right)_{0\leq\alpha,\beta\leq 3} = \mathrm{diag}(-1, 1, 1, 1)$.

The notion of simplest Lorentzian manifold can even be formalized, since by the Killing–Hopf Theorem [2, Chapter 8], Minkowski space is the unique simply connected complete Lorentzian space of vanishing constant sectional curvature. For the vacuum Einstein equations, $\mathrm{Ric}(g) = 0$, it should therefore be regarded as the trivial solution, in a similar way that the scalar function $\phi := 0$ is the trivial solution to a non-linear homogeneous wave equation of the type $\Box\phi := -\partial_t^2\phi + \Delta\phi = F(\phi, \partial\phi, \partial^2\phi)$, for some non-linearity F vanishing at $(0, 0, 0)$.

The study of the stability properties of the Minkowski space is natural, both from a physical or a mathematical point of view, once the Einstein equations are understood as a system of evolution equations; this is the problem of understanding the dynamics in a neighborhood of

an equilibrium, in fact, of the simplest equilibrium of General Relativity. The subject was also clearly of interest to Choquet–Bruhat and, as for many other mathematical problems of General Relativity, she produced several of the earlier works attempting to address it. In fact, the well-posedness statement of Choquet–Bruhat on the Einstein equations[1] already provides Cauchy stability, hence the local in time stability of the Minkowski space, as long as initial data close to Minkowski space solving the constraint equations could be constructed. This topic she addressed in [5] and [6], where she also mentioned as an open problem the question of long time evolution of such data:

> The important question whether any asymptotically flat excitations of the gravitational field remain nonsingular for all time is still open.

This question will of course only be fully resolved through the monumental work of Christodoulou–Klainerman, completed in 1989 and published in 1993 in [7] establishing the global-in-time stability of the Minkowski space.

Theorem 1.1 ([7]). *Under a global smallness assumption, any strongly asymptotically flat, maximal, initial data leads to a unique globally hyperbolic regular solution to the Einstein equations which is geodesically complete and globally asymptotically flat.*

The initial value problem for the Einstein equations, including the relevant notion of initial data are reviewed below in Section 2.2, while we recall the maximal foliation in Section 4.3. On top of constructing a global solution, detailed asymptotics for the solution were obtained in [7], such as decay estimates for the curvature or for the second fundamental form of the maximal foliation used.

Prior to [7], a work of Friedrich [8] had already established a form of future stability for hyperboloidal data close to Minkowski, but at the time of writing, it was not known how to reach the required hyperboloidal data from asymptotically flat initial data. The gluing constructions of Corvino [9], Chrusciel–Delay [10] and Corvino–Schoen [11] were eventually able to produce such data, bypassing the issue of understanding the dynamics of asymptotically flat data in a neighborhood of spatial infinity and the construction of a small piece of null infinity emanating from it, the *exterior problem*. Nonetheless, since the interior dynamics are heavily influenced by the asymptotics of the data, it is clear that one should aim to address the stability problem for the larger possible class of data, and in fact, several interesting physical phenomena can only be captured by data which are not strongly localized[2].

The work of Christodoulou–Klainerman [7] not only did provide the first complete proof of the stability of Minkowski space, but thanks to the precise asymptotics of the solutions obtained, also did contain the first rigorous derivation of the laws of gravitational radiation, such as the *Bondi mass loss formula* [13]. The non-linear *memory effect* of Christodoulou [14] was also obtained essentially as an outgrowth of the estimates obtained in the proof.

The stability of Minkowski space also constitutes a landmark in the field as the first proof in this domain of research which was too long to be published as a standard journal article; making harder the task of summarizing its content and highlighting its key ideas[3]; and nowadays one could say that it has become a sort of rite of passage to study or revisit (an excerpt of) the proof or write some form of extension.

---

[1] In particular, the estimates of [3, Chapter III] leading to a reformulation of the existence theory into a fix point argument implicitly address an analysis of the difference of solutions. See also [4] for a subsequent treatment in Sobolev spaces which focuses on Cauchy stability and states the application to the construction of near Minkowskian spacetimes for finite time.

[2] See for instance [12] for historical details and a presentation of the physical relevance of slowly decaying asymptotic flat initial data.

[3] See nonetheless [15] for an introduction and background materials.

Starting in the late 70's, the study of the long-time behaviour of non-linear wave equations with small data had seen major development[4] and in particular, the standard overall strategy based on energy and decay estimates, wrapped up in a bootstrap argument, had already seen many applications for global or long-time existence. Nonetheless, using this in the context of the Einstein equations was still challenging a priori. Choquet–Bruhat had in particular identified in [17] a form of instability in wave coordinates, essentially proving that the components of the metric in wave coordinates near Minkowski cannot all be uniformly bounded in the standard linear spaces, such as the energy space. Thus, new important building blocks were needed, in particular

- A robust method to derive decay estimates for tensorial wave equations, including improved decay for certain components or derivatives of the tensor fields.
- A way to bypass the instability highlighted by Choquet–Bruhat. In [7], a key component to avoid the instability is a novel geometric framework which is adapted to the solution: the vector fields used for commutation and the foliation are not fixed a priori but constructed throughout the proof.

The first item above was initially developed in [18], following the development of commutation vector field methods for wave equations [19] as well as the understanding of the null condition [20,21]. The failure of the standard null condition is precisely what drove the instability mechanism identified by Choquet–Bruhat and thus, one of the aims of the second item was to provide a better framework to capture the null structural properties of the Einstein equations. Eventually, Lindblad–Rodnianski [22] identified that the Einstein equations in wave coordinates actually verify a weaker form of the null condition, that allows global existence and decay estimate, while the main effects of the instability of Choquet–Bruhat remain only as a small (logarithmic) growth for the worse component of the metric [23,24].

The aim of the present article is to review some ideas and context behind the stability of the Minkowski space as well as to present some recent developments concerning the study of the long time dynamics for the Einstein equations or, more generally, geometric hyperbolic partial differential equations, which uses or extend some important ideas of the proof.

1.1. *Overwiew*

In Section 2, we briefly review standard material concerning the Einstein equations. Besides fixing some notations, introducing the evolution problem in General Relativity and emphasizing the importance of gauge conditions for the Einstein equations, we also aim here to explain why the approach used for instance for orbital type stability results in the case of critical or subcritical non-linear Schrödinger or wave equations does not apply here. In Section 3, we review the problem of global existence for quasi-linear wave equations, provide a quick overview of the original Klainerman vector field method for scalar fields and the null condition in 3 spatial dimensions and end the section with a very brief discussion of the weak null condition and the approach of Lindblad–Rodnianski. Section 4 is devoted to the presentation of the work of Christodolou–Klainerman, focusing on geometric aspects. Though nowdays, there are several aspects of the proof that one would treat differently, we have chosen to focus on the original arguments. In Section 5, we review several extensions and developments on the theme of the stability of Minkowski space and we conclude in Section 6 with some works which were influenced by the analysis on the stability of Minkowski space but address other kinds of problems.

[4] See for instance [16] for a pertinent review.

1.2. *Convention*

We use standard Einstein summation convention. Greek indices $\alpha, \beta, \ldots$ are used for tensors defined on spacetimes. Latin indices i, j, k are used for tensors on spacelike hypersurfaces or the manifold of the initial data. The convention for the signature of Lorentzian metric is $- + + \ldots$. If $g_{\alpha\beta}$ denotes the components of a metric g, $g^{\alpha\beta}$ denotes the components of the inverse metric and we also used standard raising and lowering of indices on tensors throughout the article. Partial derivatives will be denoted either as $\partial_\alpha \psi$ or $\psi_{,\alpha}$. Starting from Section 4, we follow the convention of [7] that bold quantities such as $\mathbf{R}$, $\mathbf{D}$ are associated to the spacetime geometry and thus denotes the spacetime Riemann tensor and the Levi-Civita connexion, while R, ∇ refers to the geometry of a spacelike slice.

The article is written assuming standard knowledge of differential and Riemannian geometry as well as partial differential equations, but does not assume a priori knowledge concerning General Relativity or the Einstein equations.

2. Background material

2.1. *The Einstein equations and their hyperbolicity*

General Relativity can be understood as the study of 4-dimensional[5] Lorentzian manifolds $(\mathcal{M}, g)$ solving the Einstein equations

$$\mathrm{Ric}(g) - \tfrac{1}{2}\mathrm{S}(g) \cdot g + \Lambda g = T, \tag{2.1}$$

where $\mathrm{Ric}(g)$ and $\mathrm{S}(g)$ denotes respectively the Ricci and scalar curvature of the Lorentzian metric g, $\Lambda \in \mathbb{R}$ is the so-called cosmological constant and T is a source term, arising from matter fields, the so-called energy–momentum tensor. This source term, T, should not be considered as pure external data since the evolution of matter fields, whether they are constituted of electromagnetism fields, fluids, particles, ... are themselves subject to an evolution equation depending the metric g. Thus, unless $T = 0$, the Einstein equations need to be completed by the evolution equations for the matter to get a closed system. When $T = 0$, Equation (2.1) reduces to $\mathrm{Ric}(g) = \Lambda g$, the *vacuum Einstein equations*. Minkowski space is the trivial solution to (2.2) with $\Lambda = 0$, while the de-Sitter and Anti-de-Sitter space solve the equation with $\Lambda > 0$ and $\Lambda < 0$ respectively. In the rest of the article[6], we thus focus on $\Lambda = 0$ together the vacuum condition $T = 0$, in which case the equations reduce to the Ricci flat condition

$$\mathrm{Ric}(g) = 0. \tag{2.2}$$

Perhaps the most important property of the Einstein equations is that they are geometric: given $\phi : \mathcal{M} \to \mathcal{M}$ a diffeomorphism, g solves (2.2) if and only if $\phi^\star g$ solves (2.2), where $\phi^\star g$ denotes the pullback of the metric g by ϕ.

This already implies that one cannot expect to solve the Einstein equations based purely on a traditional, direct, application of the methods from partial differential equations. Indeed, those typically gives existence and uniqueness at the same time and here uniqueness can only hold in the geometric sense, that is to say up to the above diffeomorphism invariance.

More directly, the Ricci tensor of a metric g, viewed as a (non-linear) partial differential operator on the components of g, takes the form

$$\mathrm{Ric}(g)_{\alpha\beta} = -\tfrac{1}{2}\Box_g g_{\alpha\beta} + \partial_\alpha \Gamma_\beta + \partial_\beta \Gamma_\alpha + Q_{\alpha\beta}(\partial g, \partial g), \tag{2.3}$$

where

[5] The theory can easily be written in other dimensions, but on top of being the physical dimension, dimension 4 plays a distinguished role for the problem of the stability of Minkowski space, cf. Section 3.4 below.

[6] Apart from Section 5.4, when we will review results concerning $T \neq 0$.

- $\Box_g$ is the (scalar) wave operator:
$$\Box_g \psi := \frac{1}{\sqrt{|\det g|}} \left(\partial_\alpha \sqrt{|\det g|} g^{\alpha\beta} \partial_\beta \psi \right).$$
- Γ_β are traces of the Christoffel symbols $\Gamma^\rho_{\alpha\gamma}$:
$$\Gamma_\beta := g_{\beta\rho} g^{\alpha\gamma} \Gamma^\rho_{\alpha\gamma} := g_{\beta\rho} g^{\alpha\gamma} \tfrac{1}{2} g^{\rho\delta} (g_{\delta\gamma,\alpha} + g_{\alpha\delta,\gamma} - g_{\alpha\gamma,\delta}).$$
- $Q_{\alpha\beta}(\partial g, \partial g)$ denotes a quadratic form in the first derivatives of g, with coefficients depending on g.

For a fixed metric g, $\Box_g$ is a second-order hyperbolic operator, but the terms $\partial_\alpha \Gamma_\beta + \partial_\beta \Gamma_\alpha$ contains second-order derivatives of g and thus a priori spoil the hyperbolicity of $\mathrm{Ric}(g)$. To retrieve the hyperbolicity, one needs to address the geometric invariance mentioned above, for instance by considering the equations in a particular coordinate system. The standard one, which of course was used by Choquet–Bruhat in her local well-posedness proof [3], is given by the so-callled *wave coordinates*.

Definition 1. A local coordinate system $\left(\mathscr{U} \subset \mathscr{M}, \phi := (x^\alpha)_{\alpha=0,1,2,3} : \mathscr{U} \to \mathbb{R}^4\right)$ of a Lorentzian manifold $(\mathscr{M}, g)$ is said to consist of *wave coordinates* if, $\forall \alpha = 0, \ldots, 3$, the function x^α verifies the wave equation
$$\Box_g x^\alpha = 0. \tag{2.4}$$

Since in general, one has $\Box_g x^\alpha = -g^{\gamma\beta} \Gamma^\alpha_{\gamma\beta} = -g^{\beta\alpha} \Gamma_\beta$, we have as a consequence:

Lemma 2.1. *In wave coordinates, the vacuum Einstein equations (2.2) reduces to a system of the form*
$$\Box_g g_{\alpha\beta} = Q_{\alpha\beta}(\partial g, \partial g), \tag{2.5}$$

The Equations (2.5) are a system of *quasi-linear* wave equations. For such equations, standard energy methods can be applied to prove local existence[7] in high regularity classes.

Another important property, which again can be seen to arise from the geometric invariance, is that (2.5) are scale invariant: if $g_{\alpha\beta}(x^\alpha)$ solves (2.5), restricting appropriately the domain of definition, so does ${}^\lambda g(x) := g_{\alpha\beta}(\lambda x^\alpha)$, for any $\lambda > 0$. In three spatial dimensions, the Sobolev space left invariant by this scaling is the homogeneous space $\mathring{H}^{3/2}$. In particular, this implies that one cannot hope to study the Einstein equations, even locally and thus a fortiori globally in time, without controlling more than one derivative of the metric components, for instance by taking derivatives of the equations.

2.2. *The constraint equations and the initial value problem*

Given any spacelike hypersurface Σ, the Gauss–Codazzi equations implies, for any solution of (2.2), a system of constraint equations for the first and second fundamental forms h, k associated to Σ,
$$\mathrm{S}(h) + (\mathrm{tr}\, k)^2 - |k|^2 = 0, \tag{2.6}$$
$$\mathrm{div}\, k - \mathrm{d}(\mathrm{tr}\, k) = 0, \tag{2.7}$$
where $\mathrm{S}(h)$ denotes the scalar curvature of the Riemannian metric h, and operators and traces are taken with respect to h.

The initial value problem in General Relativity then consists, given a Riemannian manifold (Σ, h) and a 2-tensor k verifying (2.6)–(2.7), in the construction of a Lorentzian manifold $(\mathscr{M}, g)$

[7] See for instance [25, Chapter 1.4].

solving (2.2) together with an embedding $i : \Sigma \hookrightarrow \mathscr{M}$ such that (h, k) coincides with the first and second fundamental forms of i. Such an $(\mathscr{M}, g)$ is then called a *development* of the data (Σ, h, k).

The Choquet–Bruhat theorem [3], completed by the Choquet—Bruhat–Geroch construction of the maximal solution[8] [27], then provides, given any initial data set (Σ, h, k), the existence and uniqueness, up to the diffeomorphism invariance, of a *maximal globally hyperbolic*[9] development.

2.3. *Asymptotic flatness*

The Euclidean data $(\mathbb{R}^3, \delta, 0)$ form of course a solution to (2.6)–(2.7) and its maximal Cauchy development is given by the Minkowski spacetime. For any concept of perturbations of the Minkowski spacetime to make sense, one thus also needs to understand how to construct initial data sets which are arbitrary close to $(\mathbb{R}^3, \delta, 0)$. In particular, we would like the pertubations from the Euclidean data to be *localized*, i.e. to tend to 0 near infinity. Physically, this corresponds to *asymptotically flat* initial data sets and gives rise to spacetimes describing *isolated systems*. From the point of view of analysis, this localization is important, as it is needed for the derivation of decay estimates, which are at the core of the stability proofs. To introduce this quantitatively, we consider initial data sets (Σ, h, k) such that Σ possesses an asymptotic end, that is, the complement of a compact set of Σ is diffeomorphic to the complement of a ball in $\mathbb{R}^3$. This allows one to introduce a coordinate system (x^1, x^2, x^3) in a neighborhood of infinity $r := \sqrt{\sum_{i=1}^3 (x^i)^2} \to +\infty$. The original proof [7] then applies to *strongly asymptotically flat initial data*, which by definition verifies the following expansion[10]

$$h_{ij} = (1 + 2M/r)\delta_{ij} + o_4(r^{-3/2}), \tag{2.8}$$

$$k_{ij} = o_3(r^{-5/2}), \tag{2.9}$$

for a non-negative number $M \geq 0$. The positive mass-theorem [28,29] in fact guarantees that for such data, the mass M is strictly positive unless the initial data is flat. In particular, one cannot consider compactly supported perturbations of the flat data. The construction of asymptotically flat initial data near the Euclidean data was first considered in [5] and [6]. The problem of producing initial data with arbitrary decay has been more recently addressed in [30] and [31]. We refer to [32, Chapter 3] for a more precise introduction to asymptotic flatness in General Relativity.

3. Global existence for quasi-linear wave equations and the basic vector field method

3.1. *A strategy for quasi-linear wave equations*

In view of the form of the reduced Equations (2.5), it is natural to first think of the stability of the Minkowski spacetime as a specific case of global existence for quasi-linear wave equations of the form

$$(-\partial_t^2 + \Delta) u_i + F^{\alpha\beta}(u)\partial_\alpha\partial_\beta(u_i) = Q_i(\partial u, \partial u), \tag{3.1}$$

$$u(t = 0) = u_0, \tag{3.2}$$

$$\partial_t u(t = 0) = u_1, \tag{3.3}$$

[8] See also the modern treatment by Sbierski [26] for an improved construction of the maximal solution, which contrary to [27] does not rely on Zorn's lemma.

[9] Here by *globally hyperbolic*, we mean the existence of a *Cauchy hypersurface*, that is an achronal hypersurface such that any inextendible causal curve register once and only once on this hypersurface, cf. [2, Chapter 14].

[10] Recall that $f = o_m(r^{-k})$ if $\partial_l f = o(r^{-k-|l|})$, for any multi-index l of length $|l| \leq m$.

for an unknown $u:(t,x)\to(u_i)(t,x)\in\mathbb{R}^N$, non-linearities such that $F^{\alpha\beta}(u)=F^{\beta\alpha}(u)$ are smooth functions of u vanishing at 0, $Q_i(\partial u,\partial u)$ which are quadratic forms in the first derivatives of u and initial data (u_0,u_1) which are small, sufficiently regular and localized.

By now, the standard approach for this type of problems is essentially based on

- A scale of weighted energy spaces to control the data and the solutions. We denote the top order energy norm at time t of a solution u by $E_N[u(t)]$, the index N referring to the order in the scale, i.e. $E_N[u]$ controls up to N derivatives of u in a weighted L^2-type space.
- Energy estimates allowing to propagate global bounds of the solutions, in particular at top order of regularity, of the form
$$E_N[u(t)]\le CE_N[u(0)]+\int_0^t \mathrm{NL}(s)\,\mathrm{d}s, \tag{3.4}$$
where NL denotes error terms arising from the non-linearities and $C>0$ is independent of u.
- Decay estimates that allow to control the non-linearities, leading, for instance, to an estimate of the form
$$|\mathrm{NL}(s)|\lesssim\frac{E_N[u(s)]}{(1+s)^q}, \tag{3.5}$$
with $q>0$.
- A bootstrap argument: by local existence, for initial data such that $E_N[u(0)]\le\epsilon$, a solution u exists on some interval $[0,T]$, for some $T>0$, and such that $E_N[u(t)]\le 2C\epsilon$, for all $t\in[0,T]$. Assuming $q>1$ for the time being in (3.5), a Grönwall argument combined with the energy estimate (3.4) and the decay estimate (3.5) then allows to improve the bootstrap assumption to $E_N[u(t)]\le C\epsilon+O(\epsilon^2)\le 3/2C\epsilon$. By a standard continuity argument, it then follows that the solution exists for all times and verifies, for all $t\ge 0$,
$$E_N[u(t)]\le C\epsilon+O(\epsilon^2).$$

To be more specific and check under which conditions, one can obtain the estimate (3.5) with $q>1$, we recall below the standard vector field method, as introduced in [19], that provides both a first definition of the spaces E_N and the decay estimate (3.5).

3.2. *The vector field method of Klainerman*

We consider the wave operator $\Box:=-\partial_t^2+\Delta$, where Δ is the usual Laplacian in $\mathbb{R}^n$ and recall first the basic energy estimate.

For ϕ and S defined on $\mathbb{R}_t\times\mathbb{R}^n_x$, decaying sufficiently fast in x and solving $\Box\phi=S$, the energy
$$E[\phi](t):=\int_{x\in\mathbb{R}^n}(\partial_t\phi(t,x))^2+|\nabla_x\phi(t,x)|^2\,\mathrm{d}x$$
verifies for all t,
$$E[\phi](t)=E[\phi](0)+\int_0^t\int_{\mathbb{R}^n}|\partial\phi\cdot S|\,\mathrm{d}x\,\mathrm{d}s,$$
which one can (at least formally) derive by multiplying the wave equation by $\partial_t\phi$ and integration by parts. Moreover, if ϕ solves instead the perturbed wave equation
$$(-\partial_t^2+\Delta)\phi+F^{\alpha\beta}\partial_\alpha\partial_\beta\phi=S,$$
with $F^{\alpha\beta}=F^{\beta\alpha}$ sufficiently small ($\|F\|_{L^\infty_{t,x}}<1/4$), then the above is replaced by the energy estimate:
$$E[\phi](t)\le 2E[\phi](0)+2\int_0^t\int_{\mathbb{R}^n}|\partial\phi|\cdot(|S|+|\partial\phi||\partial F|)\,\mathrm{d}x\,\mathrm{d}s. \tag{3.6}$$

To construct higher order norms that are still propagated by the equation, one then considers vector fields that commute with the linear wave operator:

Lemma 3.1. *Let Z be any of the vector fields*

- *Translations* $\partial_t, \partial_{x^i}$,
- *Rotations* $\Omega_{ij} = x^i\partial_{x^j} - x^j\partial_{x^i}$,
- *Hyperbolic rotations* $\Omega_{0i} = t\partial_{x^i} + x^i\partial_t$.
- *Scaling* $S = t\partial_t + \sum_i x^i\partial_{x^i}$.

Then, $[\Box, Z] = 0$, *apart from the scaling which verifies* $[\Box, S] = 2\Box$.

Denote by $\mathscr{A}$ the set of all such vector fields and recall that the Z vector fields form an algebra: $[Z_1, Z_2] \in \text{span}\mathscr{A}$, for all $Z_1, Z_2 \in \mathscr{A}$. We introduce an ordering on $\mathscr{A}$ and use standard multi-index notations below. In particular, for a multi-index α, $|\alpha|$ denotes its length and Z^α denotes a differential operator of order $|\alpha|$ obtained as a composition of $|\alpha|$ vector fields within $\mathscr{A}$.

The decay estimate is then essentially incoporated in the following weighted Sobolev inequality.

Proposition 2. *Let $\psi := \psi(t,x)$ defined on $[0,T]\times\mathbb{R}^n_x$ be a sufficiently regular function. Then,* $\forall (t,x) \in [0,T]\times\mathbb{R}^n_x$,

$$|\psi(t,x)| \lesssim \frac{1}{(1+t+|x|)^{(n-1)/2}(1+|t-|x||)^{1/2}} \sum_{|\alpha|\leq (n+2)/2,\, Z^\alpha\in\mathscr{A}^{|\alpha|}} \|Z^\alpha\psi(t,.)\|_{L^2_x}. \tag{3.7}$$

Let ψ be a solution to the linear wave equation $\Box\psi = 0$. Applying (3.7) to $\partial\psi$, we obtain

$$\begin{aligned}
|\partial\psi(t,x)| &\lesssim \frac{1}{(1+t+|x|)^{(n-1)/2}(1+|t-|x||)^{1/2}} \sum_{|\alpha|\leq (n+2)/2,\, Z\in\mathscr{A}^{|\alpha|}} \|Z^\alpha\partial\psi(t,.)\|_{L^2_x} \\
&\lesssim \frac{1}{(1+t+|x|)^{(n-1)/2}(1+|t-|x||)^{1/2}} \sum_{|\alpha|\leq (n+2)/2,,\, Z\in\mathscr{A}^{|\alpha|}} \|\partial Z^\alpha\psi(t,.)\|_{L^2_x} \\
&\lesssim \frac{1}{(1+t+|x|)^{(n-1)/2}(1+|t-|x||)^{1/2}} \sum_{|\alpha|\leq (n+2)/2,,\, Z\in\mathscr{A}^{|\alpha|}} E[Z^\alpha\psi]^{1/2}(0),
\end{aligned}$$

where we have used that $\Box Z^\alpha\psi = 0$ and hence $E[Z^\alpha\psi](t) = E[Z^\alpha\psi](0)$ from energy conservation.

Thus, we have obtain a proof of *decay* for solutions to the homegeneous wave equation. Moreover, we see that a natural scale of energy norms is given by

$$E_N[u(t)] := \sum_{|\alpha|\leq N,,\, Z\in\mathscr{A}^{|\alpha|}} E[Z^\alpha u(t)].$$

Remark 3.2. We see from (3.7) that away from the the light cone region $t = |x|$, for instance, in the interior region, $|x| < (1/2)t$, the strong decay $|\partial\phi| \lesssim t^{-3/2}$ holds. Provided the initial data enjoy enough decay, this estimate can be improved[11]. One way to capture extra decay is to consider a stronger weighted energy norm that also make use of the multiplier $K_0 := 2tS + (-t^2 + |x|^2)\partial_t$. This vector field, first used by Morawetz in [33], also played an important role in the estimates of [7], cf. Section 4.4 below.

Remark 3.3. In the strategy described above, vector fields play two different kind of roles. First, the overall energy estimates can be obtained by *multiplying* the wave equation by $X(\phi)$, for $X = X^\alpha\partial_{x^\alpha}$ a vector field, and integration over suitable regions. For instance, the basic energy estimate (3.6) relies on the vector field multiplier $X = \partial_t$, while we introduce just above the multiplier $X = K_0$, leading to weighted energy estimates. The other occurrence of the vector fields is via *commutation*, cf. Lemma 3.1 and Proposition 2. The existence of these commutation vector fields is highly connected with the symmetry of the equations, hence is less likely to be useful outside from questions related to the stability of Minkowski. On the other hand, the use of multiplier vector fields has an extremely wide range of applications. In fact, it is part

[11] In fact, for compactly supported initial data, Huygens' principle ensures that the support of the solution is located in a cylindrical strip around the light-cone.

of the *general method of energy currents*, cf. [33,34], and many recent developments in General Relativity and geometric hyperbolic partial equations rely on some applications or extensions of these principles, cf. Section 5.5 below.

3.3. *Global existence with $n \geq 4$*

Returning to the scheme introduced in Section 3.1, commuting the equations N-times and applying the energy estimate (3.6), we obtain the top order energy estimate (3.4) with an error term NL which can be estimated as

$$|NL|(s) \lesssim E_N[u(s)] \cdot \|\partial^\alpha u(s)\|_{L^\infty_x},$$

with $|\alpha| \leq N/2$. If $N/2 \geq (n+2)/2$, we can then appeal to the inequality (3.7), to obtain (3.5) with $q = (n-1)/2$. In particular, with $n \geq 4$, we have $q > 1$ and we can improve the bootstrap assumptions as explained in Section 3.1, while we see that we only have $q = 1$ if $n = 3$, so that this strategy a priori fails.

3.4. *The case $n = 3$: the null condition*

With $n = 3$, the linear decay and a Grönwall argument in (3.6) a priori gives a logarithmic growth. That this growth can in fact lead to blow up for certain systems of quasi-linear wave equations was proved in the celebrated works of John [35,36]. Nonetheless, global existence can be recovered provided the non-linearities verify certain structural properties, in particular the *null condition* [21,37], which guarantees that the non-linear terms always contain at least a derivative *tangential* to the light-cone.

For instance, for the semi-linear terms Q_i in (3.1), such a condition would imply, for $t \geq 0$, that they can be written as a linear combination with bounded coefficients of products of the form

$$\underline{\partial} u \cdot \partial u,$$

where ∂ denote all possible derivatives, while $\underline{\partial} \in \{\partial_t + \partial_r, e_A, A = 1,2\}$, with e_A derivatives tangential to the spheres of constant (t,r). The relevance of these tangential derivatives is that they verify an improved decay estimate. In fact, for any tangential derivative $\underline{\partial}$, one has a decomposition of the form

$$(1+|t|+|x|)\underline{\partial} = \sum_{Z^k \in \mathscr{A}} c_k Z^k,$$

for some uniformly bounded coefficients c_k, and the Z vector fields commute with the wave operator from Lemma 3.1, the improved decay is clear. Using the extra decay, an estimate similar to (3.5) and global existence can then be derived with some minor modifications[12].

3.5. *The Einstein equations in wave gauge*

Unfortunately, as noted by Choquet–Bruhat in [17], the Einstein equations in the wave gauge fail to verify the classical null condition. In fact, she showed later that there is no generalization of the classical null condition for the Einstein equations [40], independently of the choice of coordinates.

[12] At top order, it is harder to fully exploit the null condition, so the top order norm E_N is often allowed to grow slowly in t, while lower order norms stay uniformly bounded. See for instance [38, inequality (5.17) in Chapter 5, p. 78]. Stronger top order estimates without growth can in fact also be proven using additional spacetime estimates. Those are related to the so-called *ghost weight* of Alinhac [39]. Although the equations take on a different form, the top order estimates in [7] are also free of additional growth.

The failure of the null condition for the $Q_{\alpha\beta}$ quadratic terms in (2.5) is actually relatively easy to capture. Indeed, the $Q_{\alpha\beta}$ terms can be decomposed into a sum of null forms, that verify the null condition, cubic terms and higher order terms which will decay fast, and extra P quadratic forms of the form

$$P_{\alpha\beta}(\partial h, \partial h) = \eta^{\rho\rho'}\eta^{\sigma\sigma'}\partial_\alpha h_{\rho\sigma}\partial_\beta h_{\rho'\sigma'} - \tfrac{1}{2}\eta^{\rho\rho'}\eta^{\sigma\sigma'}\partial_\alpha h_{\rho\rho'}\partial_\beta h_{\sigma\sigma'},$$

where $h = g - \eta$ is the pertubation of the metric g compared to the Minkowski metric η. The above expression for P does not verify the classical null condition, for instance P_{tt} will contain products of t-derivattives that do not factorize with a tangential derivative to the light-cone. Nonetheless, it still contains some structure, apparent in the various contractions between η and the two copies of ∂h. Moreover, from the wave gauge condition, one derive

$$\eta^{\alpha\beta}\partial_\alpha h_{\beta\mu} = \eta^{\alpha\beta}\tfrac{1}{2}\partial_\mu h_{\alpha\beta} + \text{h.o.t},$$

from which one can rewrite non-tangential derivatives of certain components of h as tangential derivatives. All together, one can show that while the system does not verify the null condition, and as a consequence, that some components of h will not behave as in the linear case and suffer a small (at least logarithmic) growth, this growth does not propagate through the whole system thanks to a *nilpotent* type structure. These weaker structural properties, called the *weak null structure* by Lindblad–Rodnianski [22], allowed them to eventually provide an alternative proof of stability of Minkowski spacetime, first for Schwarzschildean initial data in [41] and then general asymptotically flat initial data in [42].

Remark 3.4. This wave gauge approach explained above is less geometrical than the original proof of Christodoulou–Klainerman. In particular, geometric conclusions of the stability of Minkowski spacetime, such as curvatures estimates, the Bondi Mass law, or even the completeness of null infinity[13], which are rather direct consequences of the global existence and the estimates obtained in [7], require an extra analysis on top of the global existence of solutions in wave coordinates. On the other hand, the fact that global existence (from the point of view of pdes) can be obtained independently of a strong control on the geometry is somehow a testament of the flexibility of the method and can be advantageous. See Section 5 below for many developments based on either of the two strategies.

4. The Christodoulou–Klainerman geometric approach to the stability of Minkowski space

We describe in this section some aspects of the proof of [7]. This proof is arguably too complex to summarize accurately in a few pages, so we have chosen to focus this section mostly on geometric notions and the overall strategy.

4.1. *The metric and the time foliation*

As explained in Section 2, one is forced to make gauge choices to study the Einstein equations as an evolution problem and we start by introducing the one used in [7]. To this end, we consider $(\mathcal{M}, g)$, the maximal globally hyperbolic development of evolution of strongly asymptotically flat initial data posed on $\mathbb{R}^3$ as introduced in Sections 2.2 and 2.3. By global hyperbolicity, it follows that $\mathcal{M} \simeq \mathbb{R}\times\mathbb{R}^3$ and in particular, one can introduce a time function t and coordinates (x^1, x^2, x^3) relative to which the metric takes the form

$$g = -\phi^2(t,x)\,\mathrm{d}t\otimes\mathrm{d}t + g_{ij}(t,x)\,\mathrm{d}x^i\otimes\mathrm{d}x^j.$$

[13] See Chapter 17 in [7].

The function t can be chosen so that its level sets $\Sigma_t := \{p \in \mathcal{M},\ t(p) = t\}$ provide a foliation of $\mathcal{M}$, the normalized condition $\phi \to 1$ as $x \to \infty$ is verified on each Σ_t, and such that $\Sigma_0 = \{t = 0\}$ coincides with the[14] initial data set. We denote by $\mathbf{D}$ the Levi-Civita connexion of g and by ∇ the one associated to the spacelike metric g_{ij}.

Let k_{ij} denotes the component of the second fundamental form associated to the Σ_t foliation

$$k_{ij} := \frac{1}{2\phi}\partial_t g_{ij}. \tag{4.1}$$

To decompose the tensor fields, it is useful to introduced the orthonormal frame (T, e_i), defined by $T = (1/\phi)\partial_t$ while the e_i are tangential to Σ_t and Fermi propagated along T, so that

$$\mathbf{D}_T e_i = \phi^{-1}(\nabla_i \phi) T.$$

4.2. *The Weyl tensor*

The Weyl tensor associated to g is the traceless part of the curvature tensor

$$\mathbf{C}_{\alpha\beta\gamma\delta} := \mathbf{R}_{\alpha\beta\gamma\delta} - \tfrac{1}{2}(g_{\alpha\gamma}\mathbf{R}_{\beta\delta} + g_{\beta\delta}\mathbf{R}_{\alpha\gamma} - g_{\beta\gamma}\mathbf{R}_{\alpha\delta} - g_{\alpha\delta}\mathbf{R}_{\beta\gamma}) + \tfrac{1}{6}(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma})\mathbf{S}(g).$$

In particular, for a solution to the Einstein equations, the Weyl tensor coincides with the curvature tensor. It verifies the Bianchi equations,

$$\mathbf{D}_{[\epsilon}\mathbf{C}_{\alpha\beta]\gamma\delta} = 0, \tag{4.2}$$

and, on top of being traceless by construction, enjoys the same symmetry properties as the curvature tensor.

Instead of considering all the components of the Weyl tensor, it is preferable to introduce an irreducible decomposition, the *electro-magnetic* decomposition, which takes into account its symmetries. Given a vector field, here the unit normal T, one considers the tensors E and H, defined by

$$\begin{aligned} E_{\alpha\beta} &= \mathbf{C}_{\mu\alpha\nu\beta}T^\mu T^\nu, \\ H_{\alpha\beta} &= {}^\star\mathbf{C}_{\mu\alpha\nu\beta}T^\mu T^\nu, \end{aligned}$$

where ${}^\star C$ is the *Hodge* dual of C

$${}^\star\mathbf{C}_{\mu\alpha\nu\beta} = \tfrac{1}{2}\varepsilon_{\mu\alpha\rho\sigma}\mathbf{C}^{\rho\sigma}{}_{\nu\beta},$$

with $\varepsilon_{\mu\alpha\rho\sigma}$ denoting the components of the volume form.

From the symmetric tensors E and H, one can reconstruct the whole Weyl tensor and moreover, from the Bianchi equations, it follows that (E, H) verifies a symmetric hyperbolic system similar to the Maxwell equations. For instance, the evolution equation for E takes the form

$$-\phi^{-1}\partial_t E_{ab} + \operatorname{curl} H_{ab} = Q_{ab}(\nabla\phi, E, H, k),$$

where $\operatorname{curl} H_{ab} := (1/2)(\varepsilon_a^{cd}\nabla_d H_{cb} + \varepsilon_b^{cd}\nabla_d H_{ca})$ and Q_{ab} is a quadratic in $\nabla\phi, E, H, k$, which is similar to the differential Ampère–Maxwell law.

[14] More precisely, Σ_0 coincides with the image of the initial data set by the embedding introduced in the definition of a development, cf. Section 2.2.

4.3. *Maximal foliations and elliptic systems*

One important ingredient of the setup is to consider the Weyl tensor as one of the the main unknown and only recover a control on the metric via the definition of k, Equation (4.1) and the Gauss and Codazzi equation associated to the hypersurfaces Σ_t. However, one still needs to fix the foliation, that is to constrain the hypersurfaces Σ_t. Here, one considers *maximal* hypersurfaces[15], i.e. one imposes $\operatorname{tr} k = 0$ on each Σ_t. With this choice, the lapse function ϕ verifies the elliptic equation

$$\Delta\phi = |k|^2\phi.$$

Moreover, k also verifies an elliptic div-curv type system

$$\begin{aligned} \operatorname{curl} k_{ij} &= H_{ij}, \\ \operatorname{div} k &= 0, \end{aligned}$$

where H is the magnetic part of the electro-magnetic decomposition of the Weyl tensor. Thus, one sees that a good control on the Weyl tensor leads via elliptic estimates to a good control of the second fundamental form and all the metric quantities.

4.4. *The Bel–Robinson tensor and the vector field method applied to the Bianchi equations*

At the linearized level, one thus needs to understand the following problem: given a tensor field $W := W_{\alpha\beta\gamma\delta}$ that verifies all the symmetry and traceless properties of the Weyl tensor as well as the Bianchi equations

$$\partial_{[\epsilon} W_{\alpha\beta]\gamma\delta} = 0,$$

on Minkowski space, how can one prove estimates on W? In fact, to prove a stability statement, similar to the approach highlighted in Section 3.1, one needs both energy type estimates at the top order and decay estimates. Moreover, having in mind the importance of the null condition (cf. Section 3.4) and thus of structural properties of the non-linearity, we will see here, already at the level of the components of W, that different components of W (akin to different derivatives for scalar waves) enjoy different decay properties.

First, the energy estimates are proven using the so-called *Bel–Robinson* tensor Q

$$Q_{\alpha\beta\gamma\delta} := W_{\alpha\mu\beta\nu} W^{\mu\ \nu}_{\gamma\ \delta} + {}^\star W_{\alpha\mu\beta\nu}\, {}^\star W^{\mu\ \nu}_{\gamma\ \delta}.$$

The Bel–Robinson tensor plays the role of the standard energy–momentum tensor for matter fields. It is divergence free for W solving the Bianchi equations and it is coercive whenever contracted with 4 identically directed timelike vector fields.

On Minkowski space, a vector field method leading to energy estimates and weighted Sobolev inequalities, which encodes the decay of the fields, similar to the one for the scalar field presented in Section 3.2 can then be implemented. These ideas were first presented in [18]. Compared with the simplest case of a scalar field of Section 3.2, a notable difference comes the fact that, since W is a tensor field, it has different components, and the different components enjoy different decay properties. These different decay properties are related to the directions of propagation of the evolution equations, and thus to the null conditions. To capture them, it is thus necessary to decompose W into components tangential or transversal to the null cones. Let thus $L = \partial_t + \partial_r$,

[15] In particular, [7] requires the initial data to already be maximal, i.e. that $\operatorname{tr} k = 0$ on the initial surface. The existence of such complete maximal hypersurfaces in general, asymptotically flat Lorentzian manifold (under a barrier assumption) was established by Bartnik in [43].

$\underline{L} = \partial_t - \partial_r$ and complete $L, \underline{L}$ by a pair of orthonormal vector fields e_A, $A = 1, 2$, tangential to the spheres of constant (t, r). We then decompose W on the null frame $(L, \underline{L}, e_A)$:

$$\begin{aligned} \underline{\alpha}_{AB} &:= W_{A\underline{L}B\underline{L}}, \quad \alpha_{AB} := W_{ALBL}, \\ \underline{\beta}_A &:= \tfrac{1}{2} W_{A\underline{L}\underline{L}L}, \quad \beta_A := \tfrac{1}{2} W_{AL\underline{L}L}, \\ \rho &:= \tfrac{1}{4} W_{\underline{L}L\underline{L}L}, \quad \sigma := \tfrac{1}{4} \epsilon^{\alpha\beta} W_{\alpha\beta\underline{L}L}. \end{aligned}$$

Using the Morawetz vector field K_0 (cf. Remark 3.2), which rewrites in the null frame as $K_0 = (1/2)(v^2 L + u^2 \underline{L})$ and the Bel–Robinson tensor Q, one can then compute the energy density

$$Q(K_0, T, T, T) = \tfrac{1}{8} u^2 |\underline{\alpha}|^2 + \tfrac{1}{8} v^2 |\alpha|^2 + \tfrac{1}{4}(3u^2 + v^2)|\underline{\beta}|^2 + \tfrac{1}{4}(3v^2 + u^2)|\beta|^2 + \tfrac{3}{4}(u^2 + v^2)(\rho + \sigma),$$

where $u = t - r$, $v = t + r$, from which the different weights for the different components are already apparent.

4.5. *The Eikonal equation and geometrically defined vector fields*

In implementing the above estimates on an asymptotically flat spacetime which is not Minkowski space, one important difficulty arises from the long-range effect of the $1/r$ term of the initial data, cf. (2.8), the *Schwarzschildean* tail of the metric. Indeed, the geodesic null cones of the spacetime then diverge logarithmically from the corresponding ones of Minkowski space, and their shear, which vanishes in Minkowski space, also has bad asymptotic properties. To avoid these divergences from spoiling the estimates, the strategy of [7] is to adapt the vector fields which are used in the construction of the energy norms, using the geometry of the current spacetime rather than that of the fixed Minkowski space. For instance, the vector field K_0 is replaced by

$$K := \tfrac{1}{2}(v^2 e_+ + u^2 e_-), \tag{4.3}$$

where

(1) u is an *optical* function, that is to say, for $r \geq r_0/2$, it solves[16] the Eikonal equation,

$$g(\mathrm{d}u, \mathrm{d}u) = 0, \tag{4.4}$$

(2) $e_\pm$ are null vectors: $e_\pm = T \pm N$, where T is the forward unit normal to Σ_t and N is the outward unit normal, tangential to Σ_t, to the spheres $S_{t,u}$ of constant (t, u).

(3) $v = 2r - u$, where r is the area radius of the $S_{t,u}$ spheres.

Similarly, the angular momentum vector fields Ω_{ij} replacing the flat ones (cf. Lemma 3.1) are also modified, for instance, we want $\Omega_{ij}(u) = 0$, and thus since u is now constructed dynamically, so must the Ω_{ij}.

4.6. *Curvature estimates: commutation, error terms and the principle of signature*

As explained in Section 4.4, the components of the Weyl curvature tensor are estimated using the Bel–Robinson tensor. Moreover, to obtain higher order estimates as well as obtained improved decay, one also needs to estimate components of derivatives of W such as $\widehat{\mathscr{L}}_T W$ where $\widehat{\mathscr{L}}_X W$ for a vector field X is the modified Lie derivative[17] defined by

$$\begin{aligned} \widehat{\mathscr{L}}_X W_{\alpha\beta\gamma\delta} := \mathscr{L}_X W_{\alpha\beta\gamma\delta} &- \tfrac{1}{2}\Big({}^{(X)}\pi^\mu_\alpha W_{\mu\beta\gamma\delta} + {}^{(X)}\pi^\mu_\beta W_{\alpha\mu\gamma\delta} + {}^{(X)}\pi^\mu_\gamma W_{\alpha\beta\mu\delta} + {}^{(X)}\pi^\mu_\delta W_{\alpha\beta\gamma\mu} \Big) \\ &- \tfrac{3}{8} \mathrm{tr}^{(X)}\pi W_{\alpha\beta\gamma\delta}, \end{aligned}$$

[16] To avoid caustics, u is actually constructed as a solution to (4.4) in an exterior region, $r \geq r_0/2$, for some $r_0(t) > 0$ and matched to an another optical in the supplementary interior region.

[17] The point of the modified Lie derivative is that it has the property of preserving the tracefreeness of Weyl fields: $\widehat{\mathscr{L}}_X W^\gamma_{\beta\gamma\delta} = 0$.

where ${}^{(X)}\pi_{\alpha\beta} := (1/2)\mathscr{L}_X g_{\alpha\beta}$, with $\mathscr{L}_X$ the standard Lie derivative acting on tensor fields, is the deformation tensor of the vector field X.

In the case of the T vector field, commuting the Bianchi equations $\mathbf{D}^\alpha W_{\alpha\beta\gamma\delta} = 0$, one obtains

$$\mathbf{D}^\alpha(\widehat{\mathscr{L}}_T W_{\alpha\beta\gamma\delta}) = J(T;W)_{\beta\gamma\delta},$$

for some error term $J(T;W)$ which is a quadratic form in the deformation tensor of T and the components of W.

Consider the Bel–Robinson tensor $Q(\widehat{\mathscr{L}}_T W)$ associated to $\widehat{\mathscr{L}}_T W$. Because of the additional T derivatives, the initial data of $\widehat{\mathscr{L}}_T W$ decay faster than that of W, so that $Q(\widehat{\mathscr{L}}_T W)(K,K,K,T)$, where K is the Morawetz vector field, has finite integral on the initial hypersurface, despite the strong weights in the definition of K, cf. (4.3). The divergence theorem applied between the hypersurfaces Σ_t and Σ_0 leads to

$$\begin{aligned}\int_{\Sigma_t} Q(\widehat{\mathscr{L}}_T W)(K,K,K,T) = \int_{\Sigma_0} Q(\widehat{\mathscr{L}}_T W)(K,K,K,T) + \int_0^t \mathrm{d}t' \int_{\Sigma_{t'}} \phi \operatorname{div} Q(\widehat{\mathscr{L}}_T W)_{\beta\gamma\delta} K^\beta K^\gamma K^\delta \\ + \frac{3}{2}\int_0^t \mathrm{d}t' \int_{\Sigma_{t'}} \phi\, Q(\widehat{\mathscr{L}}_T W)_{\alpha\beta\gamma\delta}\, {}^{(K)}\pi^{\alpha\beta} K^\gamma K^\delta,\end{aligned}$$

where the first spacetime error term comes from commuting the Bianchi equations and thus depends on $J(T;W)$ while the second comes from the fact that the vector field K is not an exact conformal Killing field. Similar integral formulas hold for other derivatives of W with the amount of K vector fields contracted with the Bel–Robinson tensor depending on how fast the data decay. To prove energy estimates for the components of W and its derivatives, one thus needs to control the error terms using decay estimates, and to avoid logarithmic divergences, to exploit the special structure of certain products, similar to Sections 3.1 and 3.4. Each of these error terms can be expressed as a product of null components of the deformation tensors of K or T and the null components of $\widehat{\mathscr{L}}_T W$ and one needs to exploit the different decay or integrability rates of each term. Since there are many of those, a convenient, systematic, approach is given by the *principle of conservation of signature*: given any tensor tangential to the spheres $S_{t,u}$, obtained by contraction of a spacetime tensor via contraction with the null vectors e_3 or e_4 and projections, one assigns a signature defined as the difference between the total number of contractions with the outgoing null vector e_4 and the total number of contractions with the ingoing vector e_3. For instance, $\alpha_{AB} = W(e_4, e_A, e_4, e_B)$ is given the signature $s(\alpha) = +2$ and $\underline{\alpha}_{AB} = W(e_3, e_A, e_3, e_B)$ the signature $s(\underline{\alpha}) = -2$.

By the principle of conservation of signature, given any tensor field U which can be written as a multilinear form in an arbitrary number of tensors $U_1, \dots, U_p$ with coefficients depending only the metric, the signature of any null term of U, expressed in terms of the null components of $U_1, \dots, U_p$ is equal to the sum of the signatures of each term in the decomposition. As an illustration, the component Q_{AB34} of the Bel–Robinsor tensor has signature 0. By definition, Q is a quadratric form in the null curvature components of a Weyl field, and by the principle of conservation of signature, all quadratic terms must have signature 0, leaving as only option products of the form $\alpha\cdot\underline{\alpha}$, $\beta\cdot\underline{\beta}$, $\rho\cdot\rho$, $\rho\cdot\sigma$ and $\sigma\cdot\sigma$. This can used in a (quasi)-systematic to obtain the correct form of each error term.

4.7. *Bootstrap assumptions and the last slice argument*

Similar to Section 3.1, the overall strategy is based on a bootstrap argument. Thus, one assume the existence of $t^\star$ such that within $[0, t^\star]$, quantitative estimates, both pointwises and in energy

type norms holds. As an illustration, the L^2 exterior top order bootstrap assumptions on α and $\underline{\alpha}$ take the form

$$\|(1+u^2)^2 r^q \nabla^q \underline{\alpha}\|_{L^2(r\geq r_0/2)} + \|r^{q+2}\nabla^q \alpha\|_{L^2(r\geq r_0/2)} \leq \epsilon_0, \quad q=0,1,2,$$

and the aim is then to improve these bounds and all the other bootstrap assumptions.

One important twist on the bootstrap argument concerns the optical function u. The Eikonal equation is a non-linear transport equation and thus, solutions are determined by their initial data. In [7], in the exterior region, u is initialized from the future, that is to say, from the last slice $\Sigma_{t^\star}$. Indeed, fixing u on a spacelike slice is equivalent to choosing a radial foliation of the slice. If this is done on the initial slice, since u verify a transport equation, there is no reason why the $S(t,u)$ would get rounder as $t\to+\infty$ which would forbid the improvement of certain bootstrap assumptions. Instead, one initializes u on the last slice $t^\star$ by first solving for a radial foliation on $\Sigma_{t^\star}$ verifying improved bounds. In particular, the optical function ${}^{(t^\star)}u$ depends a priori on the time-slab $[0,t^\star]$. After the bootstrap assumptions are improved, the local existence provides an extended time-slab $[0,t^\star+\delta]$, $\delta>0$, and thus a new optical function ${}^{(t^\star+\delta)}u$ must be constructed with initialization on $\Sigma_{t^\star+\delta}$. In the limit, one then also needs to prove the convergence of the optical functions ${}^{(t)}u$ as $t\to+\infty$.

5. Developments and extensions of the stability of Minkowski space

We present here a survey of result which extends or complete the stability of the Minkowski space. The exposition is thematical rather than historical.

5.1. *Stability under weaker assumptions on the initial data: Bieri and Shen works*

The initial data in the original work of Christodoulou–Klainerman were assumed to be strongly asymptotically flat, cf. Section 2.3. In [44], Bieri proved the stability of the Minkowski space for initial data verifying

$$\begin{aligned} g_{ij} &= \delta_{ij} + o_{q+1}(r^{-\frac{s-1}{2}}),\\ k_{ij} &= o_{q+1}(r^{-\frac{s+1}{2}}),\end{aligned}$$

with $s=2$. The number of derivatives on the initial data required was also lower to $q=2$, compared to up to 4 derivatives in [7]. Note that under this assumption, one cannot apriori define the ADM mass, which requires $s>2$. This result has been recently improved in the work of Shen [45], which requires only $s>1$. A main difference of the analysis in [45] compared to [44] and [7] comes from the construction of the optical function, which is no longer initialized on a last slice, but rather on the initial slice and a timelike curve playing the role of the axis of symmetry. Another analytical difference comes from abandoning of the classical vector field method, in particular, the weighted norms obtained via contraction with the K vector field, which is replaced by the r^p-weighted method of Dafermos–Rodnianski [46].

5.2. *The double null gauge and the exterior problem*

With spatial dimension 3, due to the slow decay of waves near the light-cone $t=|x|$, many steps in the analysis of quasi-linear wave equations, including for the stability of Minkowski space, must treat differently an interior region of the form $r\leq(r_0/2)(t)$, from the exterior region $r\geq(r_0/2)(t)$, where $r_0(t)$ is the area-radius of the spheres of intersection between the initial cone $u=0$ and Σ_t. In the exterior region, it can be advantageous to not consider the Einstein equations in the

maximal gauge as in [7], but in the *double null gauge*, where the spacetime metric can be put under the form

$$g := -2\Omega^2(\mathrm{d}u \otimes \mathrm{d}\underline{u} + \mathrm{d}\underline{u} \otimes \mathrm{d}u) + \gamma_{ab}(\mathrm{d}\omega^a - X^a\,\mathrm{d}u) \otimes (\mathrm{d}\omega^b - X^b\,\mathrm{d}u).$$

Here, both u and $\underline{u}$ are optical functions, i.e. solutions to the Eikonal equations, while ω^a are coordinates on the spheres $S_{u,\underline{u}}$ of constants $(u, \underline{u})$. The level sets of u and $\underline{u}$, $C(u)$, $\underline{C}(\underline{u})$ are characteristic cones and thus the double null gauge is naturally adapted to the equations. In particular, it captures better the null properties of the equations and it is also advantageous when studying radiation phenomena taking place on the asymptotic null cone $C_{\underline{u}}$, $\underline{u} \to +\infty$.

The exterior problem can then be understood as follows: given an asymptotically flat initial data set (Σ, h, k), such that for some compact set K, the initial data on $\Sigma \setminus K$ verifies a smallness assumption. Let $(\mathscr{M}, g)$ be the maximal Cauchy development of (Σ, h, k), and let $\mathscr{M}_+$ be the domain of dependence of $\Sigma \setminus K$. Show that $\mathscr{M}_+$ can be foliated by cones $(C(u), \underline{C}(\underline{u}))$ such that the outgoing cones $C(u)$ are complete and such that the curvature and Ricci coefficients are controlled in a quantitative manner. For instance, for strongly asymptotically flat data, α and $\underline{\alpha}$ must verify the decay estimates

$$r^{7/2}|\alpha| \lesssim C, \quad r|u|^{5/2}|\underline{\alpha}| \lesssim C,$$

where C depends on the initial data and r is the area radius of the spheres $S_{u,\underline{u}}$.

The exterior problem was first addressed in the work of Klainerman–Nicolò [47] and was recently revisited in the work of Shen [48], which in particular generalized the class of initial data considered, as well as Hintz [49], whose work is based on generalized wave coordinates instead of the double null gauge.

The double null foliation is particularly well-adapted to the study of characteristic initial value problems, cf. [50]. Together with the maximal gauge and the wave gauge and their generalizations, it constitutes now one of the classical ways to write the Einstein equations, with numerous spectacular applications, such as [51].

As a complement to the exterior problem, let us also mention the *spacelike-characteristic* initial value problem, which one could coin the interior problem, in which the the initial data is posed on a compact 3-disk and on a future complete null hypersurface emanating from its boundary. The global stability of the Minkowski space in this context was proved by Graf [52] based on a gauge which is partially initialized on a timelike curve, as in [45].

5.3. *The* $1+2$*-dimensional case*

In [53] and [54], Huneau studied the stability of the Minkowski space with a translational symmetry. Due to the translational invariance, the data are no longer asymptotically flat as in Section 2.3 since they enjoy no decay in the direction of symmetry. Thanks to the symmetry, the equations reduce to a quasi-linear system of wave equations in $1+2$ dimensions. Since the decay of waves improves with the number of spatial dimensions, the setting is harder than the $1+3$ case. General quasi-linear wave equations verifying the null condition were first studied by Alinhac [39]. The work of Huneau must of course address all the extra difficulties coming from the Einstein equations. In the $1+3$ dimensional case, to address the presence of the Schwarzschildean tail of the metric, the term of order $1/r$ proportional to the mass, in [24], an ansatz for the perturbation was introduced to solve only for the remainder terms. In the work of Huneau, another ansatz must be introduced, since the explicit Schwarzchild solution no longer captures the asymptotics. Instead, here, the background metrics are no longer explicit and Huneau constructs them based on the (radial) Einstein–Rosen waves. Another important modification compared

to [24] concerns the usage of *dynamical, generalized* wave coordinates, where the coordinates x^α now verify inhomogeneous wave equations instead of (2.4), with an inhomogeneous term that depends on a non-trivial background metric as well as the actual metric (but not its derivatives). This allows to substract off some of worse source terms in the wave equations for the metric perturbations.

5.4. *The stability of Minkowski space for Einstein-matter systems*

5.4.1. *The Einstein equations coupled to wave equations*

One natural way to extend the stability results of the Minkowski space is to consider the Einstein equations coupled to matter. The matter models which are also governed by wave or tensorial wave equations are clear first candidates to be studied, since the additional fields can a priori be treated by the methods already developed for the Einstein equations. The Maxwell case was addressed by Zipser in [44] based on the Christodoulou–Klainerman framework and subsequently Loizelet [55] using the wave coordinates approach of Lindblad–Rodnianski. The Einstein-scalar field is actually contained in the analysis of Lindblad–Rodnianski [24], which interestingly, thanks to the possible matter fields, shows that the positivity of the ADM mass is logically independent from the stability properties of the Minkowski space. Indeed, one can produce initial data with negative mass by considering the Einstein-scalar field equations with an negative sign in front of the energy–momentum tensor in (2.1) and yet the main theorem of [24] still applies.

In [56], Speck studied the Einstein equations coupled to non-linear electromagnetism, which contains additional difficulties due to the intrinsic non-linearities of the matter equations, even on a fixed background. Recently, Kauffman and Lindblad have considered the case of the Einstein equations coupled to a charged scalar field [57]. To treat the case of non-vanishing charge, very precise estimates on certain metric coefficients are needed, that go beyond those contained in [24]. In the vacuum case, those were derived by Lindblad [58] where a full set of sharp asymptotics for the metric components is obtained in generalized wave coordinates.

5.4.2. *Coupling between massless and massive wave equations: the Einstein–Klein–Gordon system*

In the Einstein–Klein–Gordon system (EKG) system, the Einstein equations are coupled to a massive wave equation. At the linear level, the Klein–Gordon field ψ verifies the Klein–Gordon equation

$$\Box\psi + m^2\psi = 0, \tag{5.1}$$

with mass $m > 0$. Equation (5.1), while still sharing with the same principal symbol with the wave equation, does not respect the conformal symmetries of Minkowski space due to the extra mass term. As a consequence, the conformal vector fields, such as the scaling S, do not enjoy good commutation properties with the equation. Yet, an alternative vector field approach for the study of (5.1) was developed in [59], exploiting in particular a novel foliation of the forward light-cone by hyperboloidal spacelike hypersurfaces. The coupling to the Einstein equations is still non-trivial: if one were to strictly apply the methods of either [7] or [55] in (2.1), commuting vector fields would naturally land over to the energy–momentum tensor T and force us a priori to also commute the matter equations via the same vector fields. As a consequence, one is forced to also modify the analysis of the Einstein equations in a non-trivial way. The stability of the Minkowski space for the EKG system was first obtained in the work of LeFloch–Ma in [60] for Schwarschildean initial data and then in [61] for data with more general asymptotics, using an extension of the Klainerman hyperboloidal approach. An independent proof was obtained by

Ionescu–Pausader [62], which includes slowly decaying data, based on a novel framework for the study of quasi-linear wave equations, incorporating both Fourier analysis and physical space methods. In [63], Wang initiated an approach to this problem based on intrinsic hyperboloids, instead of using the hyperboloidal foliation of the background Minkowski space.

The methods developed for the study of systems of wave-Klein–Gordon equations have found further applications, such as in the work of LeFloch–Ma addressing $f(R)$ theories of gravity [64], where the Einstein equations are replaced by a fourth order system of wave-Klein–Gordon equations, or the analysis of the Kaluza–Klein spacetime by Huneau, Stingo and Wyatt on the stability of the Kaluza–Klein spacetime [65]. Here the analysis reduces to a quasi-linear system of wave equations on a product space $\mathbb{R}^{1+3} \times \mathbb{S}^1$, corresponding to $1+4$ dimensional Minkowski spacetime with one direction compactified on a circle. The resulting system can then be viewed as coupling wave equations to an infinite sequence of Klein–Gordon equations with different masses.

5.4.3. *The Einstein–Vlasov system*

Colissionless kinetic matter is a popular matter model widely used in astropphysics or cosmology. In this case, the Einstein equations are coupled to a kinetic equation, of unknown f called a *Vlasov* field, representing a distribution of particles, which is defined on a hypersurface of the tangent bundle

$$f : \mathscr{P} \to \mathbb{R}_+, \quad \mathscr{P} := \{(x,p) \in T\mathscr{M},\ g_x(p,p) = -m^2,\ p \text{ future-oriented}\}. \tag{5.2}$$

Here, m denotes the mass of the particles and the massless case $m = 0$ has different analytical properties from the massive one $m > 0$. The evolution equation for f is that it must be propagated by geodesics, so that it verifies the transport equation

$$v^\alpha \partial_{x^\alpha} f - p^\alpha p^\gamma \Gamma^\beta_{\alpha\gamma} \partial_{p^\beta} f = 0. \tag{5.3}$$

The energy–momentum tensor in (2.1) is given by the second moment of f. For instance, in the flat case $g = \eta$, $\mathscr{P}$ can be parametrized by

$$\mathbb{R}^{1+3}_x \times \mathbb{R}^3_p \to \mathscr{P} \tag{5.4}$$

$$(x,p) \to \left(x, (p^0, p^i) := \left(\sqrt{m^2 + |p|}, p\right)\right) \tag{5.5}$$

and the energy–momentum tensor of a Vlasov field can then be written as

$$T_{\alpha\beta}[f] := \int_{(p_i)\in\mathbb{R}^3} p_\alpha p_\beta f \frac{\mathrm{d}p}{p^0}.$$

A novel difficulty comes here from the fact that the Einstein equations and the kinetic one (5.3) are of quite different nature a priori, and thus, it is not clear that they can be treated by a common approach. For instance, if one tries to commute (2.1) with vector fields, it is clear that we will be forced to consider derivatives of the energy–momentum tensor, and thus, a priori of f and (5.3).

Within spherical symmetry, the nonlinear stability of the Minkowski space for the Einstein–Vlasov system was initiated by Rein and Rendall [66] in the massive case and Dafermos in the case of massless particules [67]. Outside from spherical symmetry, the massless case has been addressed first by Taylor, in which the Einstein equations are treated as in [7,51], in the double null gauge, while the analysis of the the Vlasov field relied on the construction of Jacobi fields and assumed compact support on the initial Vlasov field. Another proof was given in [68] by Bigorgne–Fajman–Joudioux–Smulevici–Thaller, in wave coordinates, in the case of non-compact support for the Vlasov field. The massive case has first been addressed in [69,70] by Fajman–Joudioux–Smulevici and Lindblad–Taylor and more recently by Wang in [71], using an extension of the

Fourier analysis techniques of Ionescu–Pausader to address both the Einstein and the Vlasov part of the equations.

To address the issue of the coupling between the Einstein and the Vlasov equations, a systematic, vector field approach was first developped in [72] and form the core of the strategy of [69] and [68]. Another point of interest comes from the analysis of the non-linear wave/kinetic field interactions, where structures similar to the null condition or the weak null condition have been discovered. For instance, given g a solution to the wave equation and f a Vlasov field, the non-linear terms $\partial_{x^i} g.\partial_{p_i} f$ and $[\partial_{x^i} + (v^i/\sqrt{1+|v|^2})\partial_t](g).\partial_{p_i} f$ verifies stronger estimates compared to generic products $\partial_x g.\partial_p f$, due to identities such as

$$\partial_{x^i} g \partial_{p_i} = \frac{Z(g)}{t}\partial_v - \frac{x}{t}\frac{\partial_t g}{\sqrt{1+|v|^2}}\widehat{Z} + \frac{\partial_t g}{\sqrt{1+|v|^2}}\left(S + \frac{|x|^2 - t^2}{t}\partial_t\right), \tag{5.6}$$

where Z, S (respectively $\widehat{Z}$) are vector fields commuting with the wave operator (respectively, the relativistic transport operator on Minkowski space).

5.5. *New and newer vector field methods and other approaches for stability*

After Minkowski space, one would naturally like to understand the stability properties of the other stationary solutions to the Einstein equations, which, in $3+1$ dimensions[18] are given by the Schwarzschild and Kerr family. To describe all the recent progress on this fundamental problem goes well beyond the scope of this article, but we wish only to highlight some developments concerning the approach of Section 3.

First, since these geometries have less symmetry than the Minkowski space, the commutation vector field approach to decay and the strategy presented in Sections 3.1 and 3.2 needs to be considerably altered, see Remark 3.3.

In the context of the linear scalar wave equation $\Box_{g_0}\psi = 0$ on a stationary asymptotically flat geometry, such as the Kerr or Schwarzshild spacetime, a first important new tool is the derivation of local integrated decay estimates, which in simple setting can be stated as

$$\int_0^\infty \int_{r\leq R} \mathscr{E}_{r\leq R,\chi} \lesssim \mathscr{E}(0), \tag{5.7}$$

where $\mathscr{E}_{r\leq R,\chi}(\tau)$ denotes a degenerate energy flux supported in a compact region $r \leq R$ which controls the first derivative of ψ in L^2 multiplied by a function $0 \leq \chi(r) \leq 1$ and $\mathscr{E}(0)$ denotes the initial energy of the solution. The point of the function χ is that it degenerates in the so-called *trapped region*, the physical space location of null geodesics which never escapes to infinity or fall in the black hole in the case of Schwarzschild and Kerr. Starting from an estimate of the form (5.7), one can then try to derive pointwise estimates. The r^p-weighted vector field [46], mentioned already in Section 5.1, provides now a systematic and robust way to derive these pointwise estimates. In particular, one advantage compared to previous methods is that it avoids completely the use of commutator or multiplier vector fields having t weights.

In the recent [74] by Holzegel–Dafermos–Rodnianski–Taylor, yet a new and more robust strategy for the stability of stationary states has been developed. The main feature is that the analysis is essentially localized in dyadic time slabs and the fact that two main ingredients for stability are a "black box" linear inhomogeneous energy estimate on exactly stationary metrics, together with a novel physical space top order identity that need not capture the structure of trapping and is robust to perturbation. On Schwarschild and slowly rotating Kerr, this estimate

[18] Together with the Minkowski space, it is indeed conjectured that these are the only stationary solutions of the vacuum Einstein equations [73] with $\Lambda = 0$ while the picture in higher dimension or with non-vanishing Λ is considerably more complex.

relies essentially on the red-shift effect [75]. For a more precise description of the development of this new strategy (cf. the work of Mavrogiannis [76]), see the introduction of [74].

5.6. *Development of microlocal analysis and other tools for stability problems*

The black hole geometries have all in common the presence of trapped rays. In the high frequency approximation, waves can be localized arbitrarily close to these specific geodesics cf. [77], a clear obstacle to decay, which is then linked to the size and stability properties of the trapped set[19]. One way to capture this analytically is via microlocal and semi-classical analysis, and thus, many recent progress on the stability properties of black holes relies on frequency decomposition and pseudo-differential calculus[20].

Interestingly, at least in slowly rotating Kerr, physical space techniques, exploiting the existence of a Killing tensor field (replacing the role of Killing vector fields), can still be used to prove decay, cf. Andersson–Blue [80], and in fact, the work of Giorgi–Klainerman–Szeftel [81], o ne of the pieces in the recent proof of non-linear stability of slowly rotating Kerr by Klainerman and Szeftel [82], relied on the Andersson–Blue approach.

Beyond pseudo-differential calculus, the *b-calculus* of Melrose has found many applications in general relativity. For instance, in [83], Hintz and Vasy have given a novel proof of the stability of Minkowski space, based on generalized wave coordinates, the weak null structure of Lindblad–Rodnianski and the b-calculus, obtaining as a by-product of these techniques polyhomogeneous expansions of the metric components. The proof of the stability of slowly Kerr–Sitter spacetimes [84] by Hintz–Vasy also relies heavily on such methods.

6. Conclusions: beyond stability in relativity and small data problems

The methods introduced in [7] are now the basic blocks of a large literature concerning evolution problems in hyperbolic geometric partial differential equations, including many works which go beyond the problem of stability of stationary states. For instance, the use of curvature estimates, of the null structure equations and the construction and analysis of optical functions are essential in the other two monumental works of Christodoulou, the formation of trapped surfaces [51] and the formation of shocks [85].

It is also interesting to note that the null condition, which is a key element of the small data theory of quasi-linear waves, as explained in Section 3.4 is also a key structural property for many low-regularity well-posedness issues, starting from the null form estimates of Machedon, [86], to the resolution of the bounded L^2 curvature conjecture by Klainerman–Rodnianski–Szeftel [87]. In fact, the Bel–Robinson tensor is also used in [87] to estimate the curvature, and the null structure[21] which allowed in [7] to control the error terms via decay is now used to prove bilinear and trilinear estimates with improved regularity. Finally, we also remark that [87] also relies on the maximal foliation.

The Lindblad–Rodnianski approach has also had numerous developements. On top of those mentioned in the previous section, which concerns stability problems of stationary solutions,

[19] In particular, the trapped set of Schwarschild or Kerr is *normally hyperbolic*, cf. [78] for a recent analysis on general asymptotically flat spacetimes.

[20] We note also that local integrated decay estimates in this setting are typically replaced by resolvent estimates. See for instance [79] for a recent application of microlocal analysis to the Teukolsky equation on Kerr and the analysis of the resolvent.

[21] Many more ingredients are of course used in [87] to exploit the null condition, such as a quasi-linear Yang–Mills type formulation of the equations.

in [88], Luk–Oh have addresed the stability of large dispersive solutions to the Einstein–scalar–field, which are very far from the Minkowski metric, relying in particular on the understanding of the precise asymptotic behavior of the metric components in generalized wave coordinates, as in [89]. In the recent work of Touati [90] concerning the geometric optics approximation of the Einstein equations, the nilpotent structure and the the weak null condition of the Einstein equations in generalized wave coordinates, as identified by Lindblad–Rodnianski, also plays a key role in the analysis.

In conclusion, as we hope to have convey in this survey article, the global non-linear analysis of the Minkowski space is sufficiently complex that a thorough analysis and a full proof required a deep understanding of geometric quasilinear wave equations, the Einstein equations and their fine structure. Thanks to this, the main ideas introduced in its resolution have and will continue to deeply influence many of the subsequent, important, developments in General Relativity.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

[1] H. Minkowski, "Raum und Zeit", *Jahresberichte der Deutschen Mathematiker-Vereinigung* **18** (1909), pp. 75–88. Lecture given at the 80th Assembly of German Natural Scientists and Physicians, Cologne, 21 September 1908.

[2] B. O'Neill, *Semi-Riemannian Geometry: With Applications to Relativity*, Pure and Applied Mathematics, Academic Press: New York, 1983.

[3] Y. Choquet-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), no. 1, pp. 141–225.

[4] Y. Choquet-Bruhat, "Stabilité de solutions d'équations hyperboliques non linéaires. Application à l'espace-temps de Minkovski en relativité générale", *C. R. Acad. Sci. Paris Sér. A–B* **274** (1972), pp. 843–845.

[5] Y. Choquet-Bruhat, "Stabilité initiale de l'espace temps de Minkowski", *C. R. Hebd. Séances Acad. Sci. A et B* **275** (1972), pp. 1019–1021.

[6] Y. Choquet-Bruhat and S. Deser, "On the stability of flat space", *Ann. Phys.* **81** (1973), pp. 165–178.

[7] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series, Princeton University Press: Princeton, NJ, 1993, pp. x+514.

[8] H. Friedrich, "On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure", *Commun. Math. Phys.* **107** (1986), no. 4, pp. 587–609.

[9] J. Corvino, "Scalar curvature deformation and a gluing construction for the Einstein constraint equations", *Commun. Math. Phys.* **214** (2000), no. 1, pp. 137–189.

[10] P. T. Chruściel and E. Delay, "Existence of non-trivial, vacuum, asymptotically simple spacetimes", *Class. Quant. Gravity* **19** (2002), no. 9, pp. L71–L79.

[11] J. Corvino and R. M. Schoen, "On the asymptotics for the vacuum Einstein constraint equations", *J. Differ. Geom.* **73** (2006), no. 2, pp. 185–217.

[12] L. Kehrberger, "The case against smooth null infinity IV: linearized gravity around Schwarzschild—an overview", *Phil. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **382** (2024), no. 2267, article no. 20230039.

[13] H. Bondi, "Gravitational waves in general relativity", *Nature* **186** (1960), no. 4724, p. 535.

[14] D. Christodoulou, "Nonlinear nature of gravitation and gravitational-wave experiments", *Phys. Rev. Lett.* **67** (1991), pp. 1486–1489.

[15] J.-P. Bourguignon, "Stabilité par déformation non-linéaire de la métrique de Minkowski (d'après D. Christodoulou et S. Klainerman)", *Sémin. Bourbaki* **1990/91** (1991), no. 201–203, pp. 321–358. Exp. No. 740.

[16] C. Zuily, "Solutions en grand temps d'équations d'ondes non linéaires", *Sémin. Bourbaki* **36** (1993-1994), pp. 107–144. Online at http://eudml.org/doc/110181 (accessed March 25, 2025).

[17] Y. Choquet-Bruhat, "Un théorème d'instabilité pour certaines équations hyperboliques non linéaires", *C. R. Acad. Sci. Paris Sér. A–B* **276** (1973), A281–A284.

[18] D. Christodoulou and S. Klainerman, "Asymptotic properties of linear field equations in minkowski space", *Commun. Pure Appl. Math.* **43** (1990), no. 2, pp. 137–199.

[19] S. Klainerman, "Uniform decay estimates and the Lorentz invariance of the classical wave equation", *Commun. Pure Appl. Math.* **38** (1985), no. 3, pp. 321–332.

[20] D. Christodoulou, "Global solutions of nonlinear hyperbolic equations for small initial data", *Commun. Pure Appl. Math.* **39** (1986), no. 2, pp. 267–282.

[21] S. Klainerman, "The null condition and global existence to nonlinear wave equations", in *Nonlinear Systems of Partial Differential Equations in Applied Mathematics, Part 1 (Santa Fe, N.M, 1984)*, Lectures in Applied Mathematics, American Mathematical Society: Providence, RI, 1986, pp. 293–326.

[22] H. Lindblad and I. Rodnianski, "The weak null condition for Einstein's equations", *C. R. Math. Acad. Sci. Paris* **336** (2003), no. 11, pp. 901–906.

[23] H. Lindblad and I. Rodnianski, "Global existence for the Einstein vacuum equations in wave coordinates", *Commun. Math. Phys.* **256** (2005), no. 1, pp. 43–110.

[24] H. Lindblad and I. Rodnianski, "The global stability of Minkowski space–time in harmonic gauge", *Ann. of Math. (2)* **171** (2010), no. 3, pp. 1401–1477.

[25] C. D. Sogge, *Lectures on Nonlinear Wave Equations*, Monographs in Analysis, International Press: Boston, MA, 1995, pp. vi+159.

[26] J. Sbierski, "On the existence of a maximal Cauchy development for the Einstein equations: a dezornification", *Ann. Henri Poincaré* **17** (2016), no. 2, pp. 301–329.

[27] Y. Choquet-Bruhat and R. Geroch, "Global aspects of the Cauchy problem in general relativity", *Commun. Math. Phys.* **14** (1969), no. 4, pp. 329–335.

[28] R. Schoen and S. T. Yau, "On the proof of the positive mass conjecture in general relativity", *Commun. Math. Phys.* **65** (1979), no. 1, pp. 45–76.

[29] E. Witten, "A new proof of the positive energy theorem", *Commun. Math. Phys.* **80** (1981), no. 3, pp. 381–402.

[30] A. J. Fang, J. Szeftel and A. Touati, "Initial data for Minkowski stability with arbitrary decay", preprint, 2024, 2401.14353.

[31] P. G. LeFloch and T.-C. Nguyen, "The seed-to-solution method for the Einstein constraints and the asymptotic localization problem", *J. Funct. Anal.* **285** (2023), no. 9, article no. 110106.

[32] D. Christodoulou, *Mathematical Problems of General Relativity. I*, Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS): Zürich, 2008.

[33] C. S. Morawetz, "The limiting amplitude principle", *Commun. Pure Appl. Math.* **15** (1962), no. 3, pp. 349–361.

[34] D. Christodoulou, *The Action Principle and Partial Differential Equations*, Annals of Mathematics Studies, Princeton University Press: Princeton, NJ, 2000, pp. viii+319.

[35] F. John, "Blow-up for quasilinear wave equations in three space dimensions", *Commun. Pure Appl. Math.* **34** (1981), no. 1, pp. 29–51.

[36] F. John, "Blow-up of radial solutions of $u_{tt} = c^2(u_t)\Delta u$ in three space dimensions", *Math. Appl. Comput.* **4** (1985), no. 1, pp. 3–18.

[37] D. Christodoulou, "Global solutions of nonlinear hyperbolic equations for small initial data", *Commun. Pure Appl. Math.* **39** (1986), no. 2, pp. 267–282.

[38] C. D. Sogge, *Lectures on Non-Linear Wave Equations*, 2nd edition, International Press: Boston, MA, 2008, pp. x+205.

[39] S. Alinhac, "The null condition for quasilinear wave equations in two space dimensions I", *Invent. Math.* **145** (2001), no. 3, pp. 597–618.

[40] Y. Choquet-Bruhat, "The null condition and asymptotic expansions for the Einstein equations", *Ann. Phys.* **512** (2000), no. 3–5, pp. 258–266.

[41] H. Lindblad and I. Rodnianski, "Global existence for the Einstein vacuum equations in wave coordinates", *Commun. Math. Phys.* **256** (2005), no. 1, pp. 43–110.

[42] H. Lindblad and I. Rodnianski, "The global stability of Minkowski space–time in harmonic gauge", *Ann. of Math. (2)* **171** (2010), no. 3, pp. 1401–1477.

[43] R. Bartnik, "Existence of maximal surfaces in asymptotically flat spacetimes", *Commun. Math. Phys.* **94** (1984), no. 2, pp. 155–175.

[44] L. Bieri and N. Zipser, *Extensions of the Stability Theorem of the Minkowski Space in General Relativity*, AMS/IP Studies in Advanced Mathematics, American Mathematical Society: Providence, RI; International Press: Cambridge, MA, 2009, pp. xxiv+491.

[45] D. Shen, "Global stability of Minkowski spacetime with minimal decay", preprint, 2023, 2310.07483.

[46] M. Dafermos and I. Rodnianski, "A new physical-space approach to decay for the wave equation with applications to black hole spacetimes", in *XVIth International Congress on Mathematical Physics*, World Scientific Publishing Co. Pte. Ltd.: Hackensack, NJ, 2010, pp. 421–432.

[47] S. Klainerman and F. Nicolò, *The Evolution Problem in General Relativity*, Progress in Mathematical Physics, Birkhäuser Boston, Inc.: Boston, MA, 2003, pp. xiv+385.

[48] D. Shen, "Stability of Minkowski spacetime in exterior regions", *Pure Appl. Math. Q.* **20** (2024), no. 2, pp. 757–868.

[49] P. Hintz, "Exterior stability of Minkowski space in generalized harmonic gauge", *Arch. Ration. Mech. Anal.* **247** (2023), no. 5, article no. 99.

[50] J. Luk, "On the local existence for the characteristic initial value problem in general relativity", *Int. Math. Res. Not.* **2012** (2011), no. 20, pp. 4625–4678.

[51] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society (EMS): Zürich, 2009, pp. x+589.

[52] O. Graf, "Global nonlinear stability of Minkowski space for spacelike-characteristic initial data", preprint, 2021, 2010.12434.

[53] C. Huneau, "Stability in exponential time of Minkowski space–time with a translation space-like Killing field", *Ann. PDE* **2** (2016), no. 1, article no. 7.

[54] C. Huneau, "Stability of Minkowski space–time with a translation space-like Killing field", *Ann. PDE* **4** (2018), no. 1, article no. 12.

[55] J. Loizelet, "Solutions globales des équations d'Einstein-Maxwell avec jauge harmonique et jauge de Lorentz", *C. R. Math. Acad. Sci. Paris* **342** (2006), no. 7, pp. 479–482.

[56] J. Speck, "The global stability of the Minkowski spacetime solution to the Einstein-nonlinear system in wave coordinates", *Anal. PDE* **7** (2014), no. 4, pp. 771–901.

[57] C. Kauffman and H. Lindblad, "Global stability of Minkowski space for the Einstein–Maxwell–Klein–Gordon system in generalized wave coordinates", *Ann. Henri Poincaré* **24** (2023), no. 11, pp. 3837–3919.

[58] H. Lindblad, "On the asymptotic behavior of solutions to the Einstein vacuum equations in wave coordinates", *Commun. Math. Phys.* **353** (2017), no. 1, pp. 135–184.

[59] S. Klainerman, "Remark on the asymptotic behavior of the Klein–Gordon equation in $\mathbf{R}^{n+1}$", *Commun. Pure Appl. Math.* **46** (1993), no. 2, pp. 137–144.

[60] P. G. LeFloch and Y. Ma, "The global nonlinear stability of Minkowski space for self-gravitating massive fields", *Commun. Math. Phys.* **346** (2016), no. 2, pp. 603–665.

[61] P. G. LeFloch and Y. Ma, "Nonlinear stability of self-gravitating massive fields", *Ann. PDE* **10** (2024), no. 2, article no. 16.

[62] A. D. Ionescu and B. Pausader, *The Einstein–Klein–Gordon Coupled System: Global Stability of the Minkowski Solution*, Annals of Mathematics Studies, Princeton University Press: Princeton, NJ, 2022, pp. ix+296.

[63] Q. Wang, "An intrinsic hyperboloid approach for Einstein Klein–Gordon equations", *J. Differ. Geom.* **115** (2020), no. 1, pp. 27–109.

[64] P. G. LeFloch and Y. Ma, "The Euclidean-hyperboloidal foliation method: application to $f(R)$ modified gravity", *Gen. Relat. Gravit.* **56** (2024), no. 5, article no. 66.

[65] C. Huneau, A. Stingo and Z. Wyatt, "The global stability of the Kaluza–Klein spacetime", preprint, 2023, 2307.15267.

[66] G. Rein and A. D. Rendall, "Global existence of solutions of the spherically symmetric Vlasov–Einstein system with small initial data", *Commun. Math. Phys.* **150** (1992), no. 3, pp. 561–583.

[67] M. Dafermos, "A note on the collapse of small data self-gravitating massless collisionless matter", *J. Hyperbolic Differ. Equ.* **3** (2006), no. 4, pp. 589–598.

[68] L. Bigorgne, D. Fajman, J. Joudioux, J. Smulevici and M. Thaller, "Asymptotic stability of Minkowski space–time with non-compactly supported massless Vlasov matter", *Arch. Ration. Mech. Anal.* **242** (2021), no. 1, pp. 1–147.

[69] D. Fajman, J. Joudioux and J. Smulevici, "The stability of the Minkowski space for the Einstein–Vlasov system", *Anal. PDE* **14** (2021), no. 2, pp. 425–531.

[70] H. Lindblad and M. Taylor, "Global stability of Minkowski space for the Einstein–Vlasov system in the harmonic gauge", *Arch. Ration. Mech. Anal.* **235** (2020), no. 1, pp. 517–633.

[71] X. Wang, "Global stability of the Minkowski spacetime for the Einstein–Vlasov system", preprint, 2022, 2210.00512.

[72] D. Fajman, J. Joudioux and J. Smulevici, "A vector field method for relativistic transport equations with applications", *Anal. PDE* **10** (2017), no. 7, pp. 1539–1612.

[73] P. T. Chruściel, J. L. Costa and M. Heusler, "Stationary black holes: uniqueness and beyond", *Living Rev. Relat.* **15** (2012), no. 1, article no. 7.

[74] M. Dafermos, G. Holzegel, I. Rodnianski and M. Taylor, "Quasilinear wave equations on asymptotically flat spacetimes with applications to Kerr black holes", preprint, 2022, 2212.14093.

[75] M. Dafermos and I. Rodnianski, "Lectures on black holes and linear waves", in *Evolution Equations*, Clay Mathematics Proceedings, American Mathematical Society: Providence, RI, 2013, pp. 97–205.

[76] G. Mavrogiannis, "Quasilinear wave equations on Schwarzschild–de Sitter", *Commun. Partial Differ. Equ.* **49** (2024), no. 1–2, pp. 38–87.

[77] J. Sbierski, "Characterisation of the energy of Gaussian beams on Lorentzian manifolds: with applications to black hole spacetimes", *Anal. PDE* **8** (2015), no. 6, pp. 1379–1420.

[78] P. Hintz, "Normally hyperbolic trapping on asymptotically stationary spacetimes", *Probab. Math. Phys.* **2** (2021), no. 1, pp. 71–126.

[79] P. Millet, "Optimal decay for solutions of the Teukolsky equation on the Kerr metric for the full subextremal range $|a| < M$", preprint, 2023, 2302.06946.

[80] L. Andersson and P. Blue, "Hidden symmetries and decay for the wave equation on the Kerr spacetime", *Ann. of Math. (2)* **182** (2015), no. 3, pp. 787–853.

[81] E. Giorgi, S. Klainerman and J. Szeftel, "Wave equations estimates and the nonlinear stability of slowly rotating Kerr black holes", *Pure Appl. Math. Q.* **20** (2024), no. 7, pp. 2865–3849.

[82] S. Klainerman and J. Szeftel, "Kerr stability for small angular momentum", *Pure Appl. Math. Q.* **19** (2023), no. 3, pp. 791–1678.

[83] P. Hintz and A. Vasy, "Stability of Minkowski space and polyhomogeneity of the metric", *Ann. PDE* **6** (2020), no. 1, article no. 2.

[84] P. Hintz and A. Vasy, "The global non-linear stability of the Kerr–de Sitter family of black holes", *Acta Math.* **220** (2018), no. 1, pp. 1–206.

[85] D. Christodoulou, *The Formation of Shocks in 3-Dimensional Fluids*, EMS Monographs in Mathematics, European Mathematical Society (EMS): Zürich, 2007, pp. viii+992.

[86] S. Klainerman and M. Machedon, "Space-time estimates for null forms and the local existence theorem", *Commun. Pure Appl. Math.* **46** (1993), no. 9, pp. 1221–1268.

[87] S. Klainerman, I. Rodnianski and J. Szeftel, "The bounded L^2 curvature conjecture", *Invent. Math.* **202** (2015), no. 1, pp. 91–216.

[88] J. Luk and S.-J. Oh, "Global nonlinear stability of large dispersive solutions to the Einstein equations", *Ann. Henri Poincaré* **23** (2022), no. 7, pp. 2391–2521.

[89] H. Lindblad, "On the asymptotic behavior of solutions to the Einstein vacuum equations in wave coordinates", *Commun. Math. Phys.* **353** (2017), no. 1, pp. 135–184.

[90] A. Touati, "Geometric optics approximation for the Einstein vacuum equations", *Commun. Math. Phys.* **402** (2023), no. 3, pp. 3109–3200.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 555-581
https://doi.org/10.5802/crmeca.290

Review article / *Article de synthèse*

The black hole stability problem

Le problème de la stabilité des trous noirs

Sergiu Klainerman [a]

[a] Mathematics Department, Princeton University, USA
E-mail: seri@math.princeton.edu

Abstract. This is a short survey on the black hole stability problem written in honor of Yvonne Choquet Bruhat's 100th birthday, subject very dear to her and to which she has contributed greatly, in particular her foundational local existence result [1] and the maximal globally hyperbolic development of a given initial data set. The main focus is on the recent results on the stability of slowly rotating Kerr contained in the sequence of works [2–6].

Résumé. Ce texte donne un panorama succinct du problème de la stabilité des trous noirs, rédigé en l'honneur d' Yvonne Choquet Bruhat, un sujet qui lui est cher et auquel elle a grandement contribué, notamment grâce à son résultat fondamental d'existence locale [1] et au développement maximal globalement hyperbolique d'un ensemble de données initiales. L'accent est mis principalement sur les résultats récents concernant la stabilité du Kerr en faible rotation, tels qu'exposés dans la série de travaux [2–6].

Keywords. Black holes, Stability, GCM spheres, Null frames, Non-integrable horizontal structures.

Mots-clés. Trous noirs, Stabilité, Sphères GCM, Repères nuls, Structures horizontales non intégrables.

Funding. NSF grant 220103.

Manuscript received 15 October 2024, revised 13 February 2025, accepted 14 February 2025.

1. Introduction

The Einstein Field equations (EFT) $\mathbf{R}_{\alpha\beta} - (1/2)\mathbf{g}_{\alpha\beta}\mathbf{R} = \mathbf{T}_{\alpha\beta}$ possess two remarkable family of explicit solutions:

- The Kerr family $\mathcal{K}(a,m)$, with mass $m \geq 0$ and angular momentum $J = am$, consists of stationary, asymptotically flat, solutions of the Einstein vacuum equations (EVE), $\mathbf{R}_{\alpha\beta} = 0$, that is with vanishing energy–momentum tensor $\mathbf{T} = 0$. The first example, corresponding to $a = 0$, $m > 0$, was discovered by Schwarzschild in 2016, within months after Einstein wrote the final version of his field equations. It took however a lot longer to understand its *black hole* implications. The general family was discovered by Kerr [7] in 1963.
- The Kerr–Neumann family $\mathcal{KN}(a,m,e)$, with the additional charge parameter e, consists, likewise, of stationary, asymptotically flat, solutions of the Einstein–Maxwell equations (EM) for the Einstein–Maxwell equations, corresponding to the case when $\mathbf{T}$ is the energy momentum tensor of an electromagnetic Field $\mathbf{F}$, i.e. $\mathbf{T}_{\alpha\beta} = \mathbf{F}_{\alpha\gamma}\mathbf{F}_{\beta}{}^{\gamma} - (1/4)\mathbf{g}_{\alpha\beta}\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}$. The new extended family was discovered by Neumann [8] in 1964.

Similar families of black hole solutions have been discovered in the the case the Einstein field equations with a positive cosmological constant Λ, i.e. $\mathbf{R}_{\alpha\beta} - (1/2)\mathbf{g}_{\alpha\beta}\mathbf{R} + \Lambda\mathbf{g}_{\alpha\beta} = \mathbf{T}_{\alpha\beta}$. The corresponding families of black holes are known under the names *Reissner–Nordström–de Sitter*, (corresponding to Schwarzschild), *Kerr–de Sitter* and *Kerr–Newman–de Sitter*.

In this survey I will focus on the stability of the Kerr family, considered the most important case, and make only a few references on results concerning the stability of the Kerr–Newman family in Section 5. In Section 3.5.1 I also make a few remarks on the stability of the Kerr–de Sitter family.

1.1. *Relevance of the problem*

The discovery of the Kerr black holes[1], first as explicit solutions of the Einstein Field Equations (by Roy Kerr [7]) and later as possible explanations of astrophysical phenomena has not only revolutionized our understanding of the universe, it also gave mathematicians a monumental task: to test the physical reality of these solutions. This may seem nonsensical since physics tests the reality of its objects by experiments and observations and, as such, needs mathematics to formulate the theory and make quantitative predictions, not to test it. The problem, in this case, is that black holes are by definition non-observable and thus no direct experiments are possible. Astrophysicists ascertain the presence of such objects through indirect observations[2] and numerical experiments, but both are limited in scope to the range of possible observations or the specific initial conditions in which numerical simulations are conducted. One can rigorously check that the Kerr solutions have vanishing Ricci curvature, that is, their mathematical reality is undeniable. But to be real in a physical sense, they have to satisfy certain properties which, as it turns out, can be neatly formulated in unambiguous mathematical language. Chief among them[3] is the problem of stability, that is, to show that if the precise initial data corresponding to a given Kerr are perturbed a bit, the basic features of the corresponding solutions do not change much. Indeed, if the Kerr family would be unstable under perturbations, black holes would be nothing more than mathematical artifacts, "mathematical ghosts".

1.1.1. *Final state conjecture*

The problem of stability is intimately tied the far reaching *final state conjecture* (FSC) according to which: *The evolution of general (generic), asymptotically flat, initial data sets for the Einstein vacuum field equations leads to solutions which behave, in the large, like a finite number of*

[1] According to Chandrasekhar "Black holes are macroscopic objects with masses varying from a few solar masses to millions of solar masses. To the extent that they may be considered as stationary and isolated, to that extent, they are all, every single one of them, described exactly by the Kerr solution. This is the only instance we have of an exact description of a macroscopic object. Macroscopic objects, as we see them around us, are governed by a variety of forces, derived from a variety of approximations to a variety of physical theories. In contrast, the only elements in the construction of black holes are our basic concepts of space and time. They are, thus, almost by definition, the most perfect macroscopic objects there are in the universe. And since the general theory of relativity provides a single two parameter family of solutions for their description, they are the simplest as well."

[2] The physical reality of these objects was recently put to test by LIGO-Viergo which is supposed to have detected the gravitational waves generated in the final stage of the in-spiraling of two black holes. Rainer Weiss, Barry C. Barish and Kip S. Thorne received the 2017 Nobel prize for their "decisive contributions" in this respect. The 2020 Nobel prize in Physics was awarded to R. Genzel and A. Ghez for providing observational evidence for the presence of super massive black holes in the center of our galaxy, and to R. Penrose for his theoretical foundational work: the concept of a trapped surface and the proof of his famous singularity theorem.

[3] Other such properties concern the rigidity of the Kerr family, see [9] for a current survey, or the dynamical formations of black holes from regular configurations, see the [10,11] and the introduction to [12] for an up to date account of more recent results.

Kerr black holes plus a radiative decaying term[4]. The statement, consistent with both astrophysical observations and numerical simulations, is rightly regarded as the holy grail of mathematical General Relativity and is on a par in depth, beauty and difficulty with any other major open problem in Mathematics, such as the millennium Clay problems.

Though totally out of reach in full generality, FSC is related to a myriad of deep and challenging problems which are the focus of active research. To start with the conjecture, if true, implies either that the initial data is too small to concentrate and thus perturbations must asymptotically disperse to zero (*stability of Minkowski*)[5], or that it is large enough to produce bound states, i.e. black holes (*collapse problem*)[6]. The fact that final states of sufficiently large data are described, locally, by Kerr black hole solutions requires, in particular, that any stationary solution of the equations must be a Kerr solution (*rigidity problem*[7]) and that the Kerr family is *stable* under perturbations. Moreover, singularities are expected not to appear outside black holes, at least from generic initial conditions; a statement postulating this is known as the *weak cosmic censorship* conjecture[8]. The Final State Conjecture also requires us to rigorously treat interactions of black holes such as the in-spiral colliding of black holes. This is a problem of great interest for astrophysicists and numerical relativists, yet we have close to nothing in terms of rigorous mathematical results.

2. Kerr stability

The issue of the stability of the Kerr family has been at the center of attention of GR physics and mathematical relativity for more than half a century, ever since their discovery by Kerr in [7]. In this section I recall the main features of the Kerr solutions, describe the main conjecture and the recent results which settle it in the case of small angular momentum. Sections 2.3–2.5 introduce the reader to the main difficulties of the problem, Section 2.6 to the linearized problem and Section 2.7 to the most relevant simplified model problems. These latter are then discussed in more detail in Section 3.

2.1. *Kerr spacetime*

As mentioned above, Einstein Vacuum Equations admit a remarkable two parameter family of explicit solutions, the Kerr spacetimes $\mathscr{K}(a,m)$, $0 \le |a| \le m$, depending on the parameters m (mass) and a (with $J = am$ angular momentum), which are stationary and axisymmetric. In the usual Boyer–Lindquist coordinates they take the form Let $(\mathscr{K}(a,m), \mathbf{g}_{a,m})$ denote the family of Kerr spacetimes

In Boyer–Lindquist coordinates the Kerr metric is given by

$$\mathbf{g}_{a,m} = -\frac{q^2\Delta}{\Sigma^2}(\mathrm{d}t)^2 + \frac{\Sigma^2(\sin\theta)^2}{|q|^2}\left(\mathrm{d}\phi - \frac{2amr}{\Sigma^2}\mathrm{d}t\right)^2 + \frac{|q|^2}{\Delta}(\mathrm{d}r)^2 + |q|^2(\mathrm{d}\theta)^2, \tag{1}$$

where

$$\begin{cases} \Delta = r^2 + a^2 - 2mr, \quad q = r + \mathrm{i}a\cos\theta, \\ \Sigma^2 = (r^2+a^2)|q|^2 + 2mra^2(\sin\theta)^2 = (r^2+a^2)^2 - a^2(\sin\theta)^2\Delta. \end{cases} \tag{2}$$

[4]The celebrated LIGO experimental events are understood in terms of a final state consisting of just one large, spinning black hole (Kerr black hole) plus a burst of gravitational radiation which propagated through the universe over an enormously large time before reaching us.

[5]See discussion in Section 2.4.

[6]Important progress on the related problem on the formation of trapped surfaces was recently made starting with the seminal work of Christodoulou [10]. See also the introduction to [13] for more recent results.

[7]See [9] for a survey of recent results on rigidity.

[8]See [14] and the references therein.

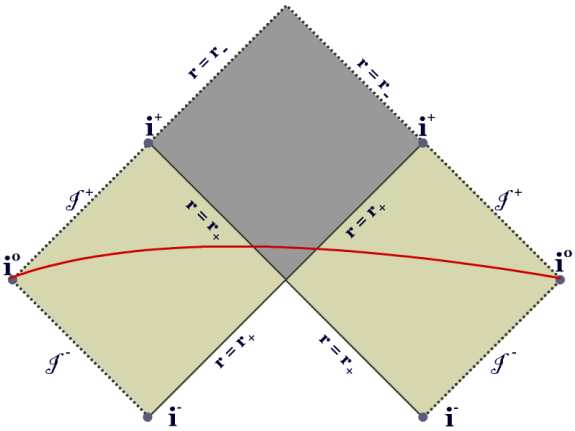


Figure 1. Penrose diagram of Kerr for $0 < |a| < m$. The surface $r = r_+$, the larger root of $\Delta = 0$, is the event horizon of the black hole, $r > r_+$ the domain of outer communication, $\mathscr{I}^+$ is the future null infinity, corresponding to $r = +\infty$.

The asymptotically flat[9] metrics $\mathbf{g}_{a,m}$ verify the Einstein vacuum equations (EVE)

$$\mathrm{Ric}(\mathbf{g}) = 0, \tag{3}$$

are stationary and axially symmetric[10], possess well-defined event horizon $r = r_+$ (the largest root of $\Delta(r) = 0$), domain of outer communication $r > r_+$ and smooth future null infinity $\mathscr{I}^+$ where $r = +\infty$. The metric can be extended smoothly inside the black hole region, see Figure 1. The boundary $r = r_-$ (the smallest root of $\Delta(r) = 0$) inside the black hole region is a Cauchy horizon across which predictability fails[11].

Here are some of the most important properties of $\mathscr{K}(a,m)$:

- $\mathscr{K}(a,m)$ possesses a canonical family of null pairs, called *principal null pairs*, of the form $(\lambda e_4, \lambda^{-1} e_3)$, with $\lambda > 0$ an arbitrary scalar function, and
$$e_4 = \frac{r^2+a^2}{|q|^2}\partial_t + \frac{\Delta}{|q|^2}\partial_r + \frac{a}{|q|^2}\partial_\phi, \quad e_3 = \frac{r^2+a^2}{\Delta}\partial_t - \partial_r + \frac{a}{\Delta}\partial_\phi. \tag{4}$$
- The horizontal structure, perpendicular to e_3, e_4, denoted $\mathscr{H}$, is spanned by the vectors
$$e_1 = \frac{1}{|q|}\partial_\theta, \quad e_2 = \frac{a\sin\theta}{|q|}\partial_t + \frac{1}{|q|\sin\theta}\partial_\phi. \tag{5}$$
The distribution generated by $\mathscr{H}$ is non-integrable for $a \neq 0$.
- The horizontal structure $(e_3, e_4, \mathscr{H})$ has the remarkable property that all components of the Riemann curvature tensor $\mathbf{R}$, decomposed relative to them, vanish with the exception of those which can be deduced from $\mathbf{R}(e_a, e_3, e_b, e_4)$, $a, b = 1, 2$.
- $\mathscr{K}(a,m)$ possesses the Killing vectorfields $\mathbf{T}, \mathbf{Z}$ which, in BL coordinates, are given by $\mathbf{T} = \partial_t, \mathbf{Z} = \partial_\phi$.
- In addition to the symmetries generated by $\mathbf{T}, \mathbf{Z}$, $\mathscr{K}(a,m)$ possesses also a non-trivial Killing tensor[12], i.e. a symmetric 2-tensor $\mathbf{C}_{\alpha\beta}$ verifying the property $\mathbf{D}_{(\gamma}\mathbf{C}_{\alpha\beta)} = 0$. The tensor carries the name of its discoverer B. Carter, see [15], who made use of it to show

[9] That is they approach the Minkowski metric for large r.

[10] That is $\mathscr{K}(a,m)$ possess two Killing vectorfields: the stationary vectorfield $\mathbf{T} = \partial_t$, which is time-like in the asymptotic region, away from the horizon, and the axial symmetric Killing field $\mathbf{Z} = \partial_\phi$.

[11] Infinitely many smooth extensions are possible beyond the boundary.

[12] Given by the expression $\mathbf{C} = -a^2\cos^2\theta\,\mathbf{g} + O$, $O = |q|^2(e_1 \otimes e_1 + e_2 \otimes e_2)$.

that the geodesic flow in Kerr is integrable. Its presence, in addition to **T** and **Z**, as a higher order symmetry, is at the heart of what Chandrasekhar, see [16], called the most striking feature of Kerr, "the separability of all the standard equations of mathematical physics in Kerr geometry".

- The Carter tensor can be used to define the Carter operator

$$\mathscr{C} = \mathbf{D}_\alpha(\mathbf{C}^{\alpha\beta}\mathbf{D}_\beta), \tag{6}$$

a second order operator which commutes with $\Box_{a,m}$. This property plays a crucial role in the proof of the stability result, Theorem 4.1, more precisely in Part II of [3].

2.2. *Main conjecture*

Conjecture. [Stability of Kerr conjecture] *Vacuum, asymptotically flat, initial data sets, sufficiently close to* $\mathscr{K}(a,m)$, $|a|/m < 1$, *initial data, have maximal developments with complete future null infinity and with domain of outer communication*[13] *which approaches (globally) a nearby Kerr solution.*

2.3. *Notions of stability*

It is instructive to consider the problem of the stability of the Kerr family in the broader context of nonlinear PDEs of the form

$$\mathscr{N}[\phi] = 0. \tag{7}$$

Consider a stationary solution ϕ_0 of (7). There are two distinct notions of stability, *orbital stability*, according to which small perturbations of ϕ_0 lead to solutions ϕ which remain close to ϕ_0 for all time, and *asymptotic stability (AS)* according to which the perturbed solutions converge as $t \to \infty$ to ϕ_0. In the case where ϕ_0 is non trivial, i.e. it depends on the spatial variables, there is a third notion, called *asymptotic orbital stability (AOS)*, to describe the fact that the perturbed solutions may converge to a different stationary solution. This happens if ϕ_0 belongs to a multi-parameter smooth family of stationary solutions, or by applying a gauge transform to ϕ_0 which keeps the equation invariant[14].

For quasilinear equations[15], such as EVE, a proof of stability means necessarily AS or AOS stability. Both require a detailed understanding of the *decay properties* of the linearized equation, i.e.

$$\mathscr{L}[\phi_0]\psi = 0, \tag{8}$$

with $\mathscr{L}[\phi_0]$ the Fréchet derivative $\mathscr{N}'[\phi_0]$. This is, essentially, a linear hyperbolic system with variable coefficients which, typically, presents instabilities[16].

In the exceptional situation, when stability can ultimately be established, one can tie all the instability modes to the following properties of the nonlinear equation:

(M1) If ϕ_λ is a family of stationary solutions, near ϕ_0, verifying $\mathscr{N}[\phi_\lambda] = 0$. Then $\psi_0 = ((\mathrm{d}/\mathrm{d}\lambda)\phi_\lambda)_{\lambda=0}$ verifies $\mathscr{N}'[\phi_0]\psi_0 = 0$, i.e. ψ_0 is a nontrivial, stationary, bound state of the linearized equations (8).

[13] This presupposes the existence of an event horizon. Note that the existence of such an event horizon can only be established a posteriori, upon the completion of the proof of the conjecture.

[14] In Kerr, both cases are present.

[15] Orbital stability can be established directly (i.e. without establishing the stronger version) only in rare occasions, such as for hamiltonian equations with weak nonlinearities.

[16] In unstable situations (8) may have exponentially growing solutions, see for example [17].

(M2) If Φ_λ is a smooth family of diffeomorphisms of the background manifold, $\Phi_0 = I$, such that $\mathcal{N}[\Phi_\lambda^*(\phi_0)] = 0$. Then $\Psi_0 = ((\mathrm{d}/\mathrm{d}\lambda)\Phi_\lambda^*(\phi_0))_{\lambda=0}$ verifies $\mathcal{N}'[\phi_0]\Psi_0 = 0$, i.e. Ψ_0 is also a stationary bound state of the linearized equation (8).

These linear instabilities are responsible for the fact that a small perturbation of the fixed stationary solution ϕ_0 may not converge to ϕ_0 but to another nearby stationary solution[17].

To prove the asymptotic convergence of ϕ to a final state ϕ_f we need to establish sufficiently strong rates of decay[18] for $\phi - \phi_f$. Rates of decay however are strongly *coordinate dependent*, i.e. dependent on the choice of the diffeomorphism (or gauge) Φ in which decay is measured. Thus, to prove a nonlinear stability result we need to know both the final state ϕ_f and the coordinate system Φ_f in which sufficient decay, and thus convergence to ϕ_f, can be established. The difficulty here is that neither ϕ_f nor Φ_f can be determined a-priori (from the initial perturbation), they have to emerge dynamically in the process of convergence. Moreover, in all examples involving nonlinear wave equations in $1+3$ dimensions, the nonlinear terms have to also cooperate, that is an appropriate version of the so called null condition has to be verified.

To summarize, given a nonlinear system $\mathcal{N}[\phi] = 0$ which possesses both a smooth family of stationary solutions ϕ_λ and a smooth family of invariant diffeomorphisms a proof of the nonlinear stability of ϕ_0 requires the following ingredients:

- The only non-decaying modes of the linearized equation $\mathcal{L}(\phi_0)\psi = 0$ are those due to the items M1–M2 above. In particular there are no exponentially growing modes.
- A dynamical construction of both the final state ϕ_f and the final gauge Φ_f in which convergence to the final state takes place.
- The nonlinear terms in the equation
$$\mathcal{L}[\phi_f]\psi = N(\psi)$$
obtained by expanding the equation $\mathcal{N}[\phi]$ near ϕ_f, in the gauge given by the diffeomorphism Φ_f, has to verify an appropriate version of the null condition.

2.4. *Stability of Minkowski space*

In the case of the Kerr family one has to deal not only with a 2-parameter family of solutions, corresponding to the parameters (a, m), but also with the entire group of diffeomorphisms[19] of $\mathcal{M}$. To understand these difficulties it helps to them with those arising in the simplest case when $a = m = 0$, i.e. stability of Minkowski space.

Until very recently the only space-time for which full nonlinear stability had been established was the Minkowski space. The first proof in [20] fits in the category of asymptotic stability, i.e. convergence back to Minkowski space, and is based on some important PDE advances of late last century:

(i) Robust approach, based on the vectorfield method, to derive quantitative decay based on generalized energy estimates and commutation with (approximate) Killing and conformal Killing vectorfields.

(ii) The *null condition* identifying the deep mechanism for nonlinear stability, i.e. the specific structure of the nonlinear terms which enables stability despite the low decay of the perturbations.

[17] The methodology of tracking this asymptotic final state, in general different from ϕ_0, is usually referred to as modulation. See for example [18,19] for how modulation theory can be used to deal with some examples of scalar nonlinear dispersive equations.

[18] To control the nonlinear terms of the equation.

[19] Indeed, according to the covariant properties of the Einstein vacuum equations we cannot distinguish between $\mathbf{g}$ and $\Phi^*\mathbf{g}$, for any diffeomorphism Φ of $\mathcal{M}$.

(iii) Elaborate bootstrap argument according to which one makes educated assumptions about the behavior of solutions to nonlinear wave equations and then proceeds, by a long sequence of a-priori estimates, to show that they are in fact satisfied. This amounts to a *conceptual linearization*, i.e. a method by which the equations become, essentially, linear[20] without actually linearizing them.

The main innovation in the proof in [20] is the choice of an appropriate gauge condition, readjusted dynamically through the convergence process, by a continuity argument, which allows one to separate the curvature estimates, treated by hyperbolic methods, from the estimates for the connection coefficients. The key point is that these latter verify transport or elliptic equations in which the curvature terms appear as sources. Thus both the curvature components and connection coefficients can be controlled by a bootstrap argument. The gauge condition is based on the constructive choice of a maximal time function t and two outgoing optical functions ${}^{(\mathrm{int})}u$[21] and ${}^{(\mathrm{ext})}u$[22] covering the interior and exterior parts of the spacetime.

Another novelty of [20] was the reliance on null frames adapted to the S-foliations induced by the level surfaces of t and u. These define integrable horizontal structures (in the language of part I of [3]), by contrast with the non-integrable ones used in the proof of Theorem 4.1 and discussed in Section 4.3.2. The functions t, u and this integrable horizontal structure can be used to define approximate Killing vectorfields used in estimating the curvature.

2.5. *Main difficulties in the Kerr stability problem*

There are a few major obstacles in passing from the stability of Minkowski to that of Kerr:

(1) The first one was already discussed in Section 2.3 in the general context of the stability of a stationary solution ϕ_0. In the case when ϕ_0 is trivial there are no nontrivial bound states for the linearized problem and thus we expect that the final state does actually coincide with ϕ_0. This is precisely the case for the special member of the Kerr family $a = m = 0$, i.e. the Minkowski space[23] $(\mathbb{R}^{1+3}, m)$. On the other hand, in perturbations of Kerr, general covariance affects the entire construction of the spacetime. In the proof of Theorem 4.1 the crucial concept of a GCM admissible spacetime is meant to deal with both finding the final parameters and the gauge in which convergence to the final state takes place.

(2) A fundamental insight in the stability of the Minkowski space was that the Bianchi identities decouple at first order from the null structure equations which allows one to control curvature first, as a Maxwell type system (see [21]), and then proceed with the rest of the solution. This cannot work for perturbations of Kerr due to the fact that some of the null components[24] of the curvature tensor are non-trivial in Kerr.

(3) Even if one succeeds in tackling the above mentioned issues, there are still major obstacles in understanding the decay properties of the solution. Indeed, when one considers the simplest, relevant, linear equation on a fixed Kerr background, i.e. the scalar wave

[20] Note that in the context of EVE, and other quasilinear hyperbolic systems, this differs substantially from the usual notion of linearization around a fixed background.

[21] The interior optical function is initialized on a timelike geodesic from the initial hypersurface.

[22] The exterior optical function ${}^{(\mathrm{ext})}u$ is initialized on the last slice $t = t_*$, by the construction of a foliation (inverse lapse foliation) initialized at space-like infinity. It is thus readjusted dynamically as $t_* \to \infty$.

[23] Note however that even though the linearized system around Minkowski does not contain instabilities, the proof of the nonlinear stability of the Minkowski space in [20] takes into account (in a fundamental way!) general covariance. Indeed the presence of the ADM mass affects the causal structure of the far, asymptotic, region of the perturbed spacetime.

[24] With respect to the *principal null directions of Kerr*, i.e. a distinguished null pair which diagonalizes the full curvature tensor, the middle component $P = \rho + i\,{}^*\rho$ is nontrivial.

equation $\Box_{a,m}\psi = 0$, one encounters serious difficulties to prove decay. Below is a very short description of these:

- *The problem of trapped null geodesics.* This concerns the existence of null geodesics[25] neither crossing the event horizon nor escaping to null infinity, along which solutions can concentrate for arbitrary long times. This leads to degenerate energy–Morawetz estimates which require a very delicate analysis.
- *The trapping properties of the horizon.* The horizon itself is ruled by null geodesics, which do not communicate with null infinity and can thus concentrate energy. This problem was solved by understanding the so called red-shift effect associated to the event horizon, which counteracts this type of trapping.
- *The problem of superradiance.* This is the failure of the stationary Killing field $\mathbf{T} = \partial_t$ to be everywhere timelike in the domain of outer communication[26], and thus, of the associated conserved energy to be positive. Note that this problem is absent in Schwarzschild and, in general, for axially symmetric solutions of EVE. In both cases however there still is a degeneracy along the horizon.
- *Superposition problem.* This is the problem of combining the estimates in the near region, close to the horizon, (including the ergoregion and trapping) with estimates in the asymptotic region, where the spacetime is close to Minkowski.

(4) Though, as seen above, the analysis of the scalar wave equation in Kerr presents formidable difficulties, it is itself just a vastly simplified model problem. A more realistic equation is the so called spin 2 wave equation, or Teukolsky equation, which presents many new challenges[27].

(5) The full linearized system, whatever its formulation, presents many additional difficulties due to its complex tensorial structure and the huge gauge covariance of the equations[28]. The crucial breakthrough in this regard is the observation, due to Teukolsky [22], that the extreme components of the linearized curvature tensor are both gauge invariant (see below in Section 2.6) and verify decoupled spin 2 equations, that is the Teukolsky equations mentioned above.

(6) A crucial simplification of the linear theory, by comparison to the full nonlinear case, is that one can separate the treatment of the gauge invariant extreme curvature components form all the other gauge invariant quantities. In the nonlinear case this separation is no longer true, all quantities need to be treated simultaneously. Moreover, methods based on separation of variables, developed to treat scalar and and spin 2 wave equations in Kerr, are incompatible with the nonlinear setting which requires, instead, robust methods to derive decay.

2.6. *Linear stability*

Linear stability for the vacuum equations is formulated in the following way. Given the Einstein tensor $\mathbf{G}_{\alpha\beta} = \mathbf{R}_{\alpha\beta} - 1/2\mathbf{R}\mathbf{g}_{\alpha\beta}$ and a stationary solution $\mathbf{g}_0$, i.e. a fixed Kerr metric, one has to solve the system of equations

$$\mathbf{G}'(\mathbf{g}_0)\,\delta\mathbf{g} = 0. \tag{9}$$

[25] In the Schwarzschild case, these geodesics are associated with the celebrated photon sphere $r = 3m$.

[26] The stationary Killing vectorfield $\mathbf{T}$ is timelike only outside of the so-called ergoregion.

[27] Unlike the scalar wave equation $\Box_{a,m}\psi = 0$, which is conservative, the Teukolsky equation is not, and we thus lack the most basic ingredient in controlling the solutions of the equation, i.e. energy estimates.

[28] As mentioned earlier, rates of decay are heavily dependent on a proper choice of gauge, thus affecting the issue of convergence.

The covariant properties of the Einstein equations, i.e. the equivalence between a solution $\mathbf{g}$ and $\Phi^*(\mathbf{g})$, leads us to identify $\delta\mathbf{g}$ with $\delta\mathbf{g} + \mathscr{L}_X(\mathbf{g}_0)$ for arbitrary vectorfields X in $\mathscr{M}$, i.e.

$$\delta\mathbf{g} \equiv \delta\mathbf{g} + \mathscr{L}_X(\mathbf{g}_0). \tag{10}$$

One can now attempt to formulate a version of linear stability for (9), loosely related to the nonlinear stability of Kerr conjecture, as follows.

Definition 2.1 (By linear stability of the Kerr metric $\mathbf{g}_0$ we understand a result which achieves the following). Given an appropriate initial data for a perturbation $\delta\mathbf{g}$, find a vectorfield X such that, after projecting away the bound states generated by the parameters a, m, according to M1–M2 in Section 2.3, a solution of the form $\delta\mathbf{g} + \mathscr{L}_X\mathbf{g}_0$ to (9), decays, relative to an appropriate null frame of $\mathscr{K}(a_0, m_0)$, sufficiently fast in time.

Remark 2.2. The definition above is necessarily vague. What is the meaning of sufficiently fast? In fact various components of the metric $\delta\mathbf{g}$, relative to the canonical null frame of $\mathscr{K}(a, m)$, are expected to decay at different polynomial rates, some of which are not even integrable. Unlike in the nonlinear context, where one needs precise rates of decay for components of the curvature tensor and Ricci coefficients (as well as their derivatives) to control the nonlinear terms, in linear theory any type of nontrivial control of solutions may be regarded as satisfactory[29]. Thus linear stability, as formulated above, can only be regarded as a vastly simplified model problem. Nevertheless the study of linear stability of the Kerr family has turned out to be useful in various ways, as we shall see below.

Historically, the following versions of linear stability have been considered.

(a) *Metric perturbations.* At the level of the metric itself, i.e. as above in (9).
(b) *Curvature perturbations.* Via the Newman–Penrose (NP) formalism, based on null frames.

The strategy followed in both cases[30] is:

- Find components of the metric (in case (a)) or curvature tensor (in case (b)), invariant with respect to linearized gauge transformations (10), which verify decoupled wave equations. The main insight of this type was the discovery, by Teukolsky [22], in the context of (b) above, that the extreme components of the linearized curvature tensor verify both these properties.
- Analyze these components by showing one of the following:
  - There are no exponentially growing modes. This is known as mode stability.
  - Boundedness for all time.
  - Decay (sufficiently fast) in time.
- Find a linearized gauge condition, i.e. a vectorfield X, such that all remaining (gauge dependent) components (at the metric or curvature level) inherit the property mentioned above: no exponentially growing modes, boundedness, or decay in time. In the physics literature this is known as the problem of reconstruction.

2.6.1. *Mode stability*

All results on the linear stability of Kerr in the physics literature during the 10–15 years after Roy Kerr's 1963 discovery, often called the "Golden Age of Black Hole Physics", are based on mode decompositions. One makes use of the separability[31] of the linearized equations, more

[29] Thus, for example, in their well known linear stability result around Schwarzschild [23], the authors derive satisfactory results (compatible with what is needed in nonlinear theory) for components of the curvature tensor, and some Ricci coefficients, but not all. Similar comments apply to [24,25].

[30] In the article we refer mainly to the curvature perturbation approach.

[31] See discussion in Section 2.1.

precisely the Teukolsky equations, on a fixed Kerr background, to derive simple ODEs for the corresponding modes. One can then show, by ingenious methods, that these modes cannot exhibit exponential growth. The most complete result of this type is due to Bernard Whiting [26].

The obvious limitation of these results are as follows:

- They are far from even establishing the boundedness of general solutions to the Teukolsky equations, let alone to establish quantitative decay for the general solutions.
- Results based on mode decompositions depend strongly on the specific symmetries of Kerr which cannot be adapted to perturbations of Kerr.

Robust methods to deal with both issues have been developed in the mathematical community, based on the vectorfield method which I discuss below.

2.6.2. *Classical vectorfield method*

The vectorfield method, as an analytic tool to derive decay, was first developed in connection with the wave equation in Minkowski space. As well known, solutions of the wave equation $\Box\psi = 0$ in the Minkowski space $\mathbb{R}^{n+1}$ both conserve energy and decay uniformly in time like $t^{-(n-1/2)}$. While conservation of energy can be established by a simple integration by parts, and is thus robust to perturbations of the Minkowski metric, decay was first derived either using the Kirchhoff formula or by Fourier methods, which are manifestly not robust. An integrated version of local energy decay, based on an inspired integration by parts argument, was first derived by Morawetz [27,28]. The first derivation of decay based on the commutations properties of $\Box$ with Killing and conformal Killing vectorfields of Minkowski space together with energy conservation appear in [29,30]. The same method also provides precise information about the decay properties of derivatives of solutions with respect to the standard null frame of Minkowski space, an important motivating factor in the discovery of the null condition [31–33].

The crucial feature of the methodology initiated by these papers, to which we refer as the classical vectorfield method, is that it can be easily adapted to perturbations of the Minkowski space. As such the method has had numerous applications to nonlinear wave equations and played an important role in the proof of the nonlinear stability of Minkowski space, as discussed in Section 2.4. It has also been applied to later versions of the stability of Minkowski in [34–42] and extensions of it to Einstein equation coupled with various matter fields in [43–48].

2.6.3. *New vectorfield method*

To derive decay estimates for solutions of wave equations on a Kerr background one has to substantially refine the classical vectorfield method. The new vectorfield method is an extension of the classical method which compensates for the lack of enough Killing and conformal Killing vectorfields in Kerr by introducing, new, cleverly designed, vectorfields whose deformation tensors have coercive properties in different regions of spacetime, not necessarily causal. The method has emerged in the last 20 years in connection to the study of boundedness and decay for the scalar wave equation in Schwarzschild and Kerr, see Section 3.2 for more details.

2.7. *Model problems*

To solve the stability of Kerr conjecture one has to deal simultaneously with all the difficulties mentioned above. This is, of course, beyond the abilities of mere humans. Instead the problem was tackled in a sequence of steps based on a variety of simplified model problems, in increasing order of difficulty. To start with we can classify model problems based on the following criteria:

(1) Whether the result refers to Schwarzschild i.e. $a = 0$, slowly rotating Kerr i.e. $|a| \ll m$ or full non-extremal Kerr $|a| < m$.

(2) Whether the result refers to linear or nonlinear stability.
(3) Whether the result, in linear theory, refers to scalar wave equation, i.e. spin 0, Teukolsky equation, i.e. spin 2, or the full linearized system.
(4) Whether the stability result, in linear theory, is a mode stability result, a boundedness result or one that establishes quantitative decay.

3. Short survey of model problems

I give below a short outline of the main developments concerning linear and nonlinear model problems for Kerr stability.

3.1. *Mode stability results*

These are mode stability results, using the method of separation of variables, obtained in the physics community roughly during the period 1963–1990. They rely on what Chandrasekhar called the most striking feature of Kerr i.e. "*the separability of all the standard equations of mathematical physics in Kerr geometry*".

(1) *Regge–Wheeler* [49]. Even before the discovery of the Kerr solution physicists were interested in the mode stability of Schwarzschild space, i.e. $\mathscr{K}(0,m)$. The first important result goes back to Regge and Wheeler [49], in which they analyzed linear, metric perturbations, of the Schwarzschild metric. They showed that in a suitable gauge, Equation (9) decouples into even-parity and odd-parity perturbations, corresponding to axial and polar perturbations. The most important discovery in that paper is that of the master Regge–Wheeler equation, a wave equation with a favorable potential, verified by an invariant scalar component ϕ of the metric, i.e.

$$\Box_m \phi = V\phi, \quad V = \frac{4}{r^2}\left(1 - \frac{2m}{r}\right). \tag{11}$$

where $\Box_m$ denotes the wave operator of the Schwarzschild metric of mass m. The R–W study was completed by Vishveshwara [50] and Zerilli [51]. A gauge-invariant formulation of *metric perturbations* was then given by Moncrief [52].

(2) *Teukolsky* [22]. The curvature perturbation approach, near Schwarzschild, based on the Newman–Penrose (NP) formalism was first undertaken by Bardeen–Press [53]. This approach was later extended to the Kerr family by Teukolsky [22], see also [54], who made the important discovery that the extreme curvature components, relative to a principal null frame, are gauge invariant and satisfy decoupled, separable, wave equations. The equations, bearing the name of Teukolsky, are roughly of the form

$$\Box_{a,m}\psi = \mathscr{L}[\psi] \tag{12}$$

where $\mathscr{L}[\psi]$ is a first order linear operator in ψ.

(3) *Chandrasekhar* [55]. In [55] Chandrasekhar initiated a transformation theory relating the two approaches. He exhibited a transformation which connects the Teukolsky equations to a Regge–Wheeler type equation. In the particular case of Schwarzschild the transformation takes the Teukolsky equation to the Regge–Wheeler equation in (11). The *Chandrasekhar transformation* was further elucidated and extended by Wald [56] and recently by Aksteiner and al [57]. Though originally it was meant only to unify the Regge–Wheeler approach with that of Teukolsky, the Chandrasekhar transformation, and various extensions of it, turn out to play a fundamental role in the stability problem.

(4) *Whiting* [26]. As mentioned before, the *full mode stability*, i.e. lack of exponentially growing modes, for the Teukolsky equation on Kerr is due to Whiting[32], see [26]. Stronger quantitative versions were proved in [61–63].

(5) *Reconstruction.* Once we know that the Teukolsky variables, i.e. the extreme components of the curvature tensor verify mode stability, i.e. there are no exponentially growing modes, it still remains to deal with the problem of reconstruction, i.e. to find a gauge relative to which all other components of the curvature and Ricci coefficients enjoy the same property. We refer the reader to Wald [56] and the references within for a treatment of this issue in the physics literature.

3.2. *Quantitative decay for the scalar wave equation*

As mentioned in Section 2.6.1, *mode stability* is far from establishing even the boundedness of solutions. To achieve that[33] and, more importantly, to derive realistic decay estimates, one needs an entirely different approach based on what one may call the "new vectorfield method", in section 2.6.3. The method has emerged in connection to the study of boundedness and decay for the scalar wave equation in $\mathscr{K}(a,m)$,

$$\Box_{a,m}\phi = 0. \tag{13}$$

The starting and most demanding part of the new method, which (in the case $a = 0$) appeared first in the pioneering work of Blue and Soffer [65], is the derivation of a global, combined *Energy–Morawetz* estimate which degenerates in the trapping region. Once an Energy–Morawetz estimate is established one can commute with the Killing vectorfields of the background manifold, and the so called red shift vectorfield introduced in [66], to derive uniform bounds for solutions. The most efficient way to also get decay, and solve the *superposition problem* (see Section 2.5), originating in [67], is based on the presence of family of r^p*-weighted*, quasi-conformal vectorfields defined in the non-causal, far r region of spacetime[34].

The first Energy–Morawetz type results for scalar wave Equation (13) in Schwarzschild, i.e. $a = 0$, are due to Blue–Soffer [65,69] and Blue–Sterbenz [70], based on a modified version of the classical Morawetz integral energy decay estimate. Further developments appear in the works of Dafermos–Rodnianski, see [66,67], and Marzuola–Metcalfe–Tataru–Tohaneanu [71]. The vectorfield method can also be extended to derive decay for axially symmetric solutions in Kerr, see [72] and[35] [73], but it is known to fail for general solutions in Kerr, see Alinhac [74].

In the absence of axial symmetry the derivation of an Energy–Morawetz estimate in $\mathscr{K}(a,m)$ for $|a/m| \ll 1$ requires a more refined analysis involving both the vectorfield method and either micro-local methods or mode decompositions. The first full quantitative decay[36] result, based on micro-local analysis techniques, is due to Tataru–Tohaneanu [76]. The derivation of such an estimate in the full sub-extremal case $|a| < m$ is even more subtle and was achieved by Dafermos–Rodnianski–Shlapentokh–Rothman [60] by combining the vectorfield method with a full separation of variables approach. A purely physical space proof of the Energy–Morawetz estimate for small $|a/m|$, which avoids both micro-local analysis and mode decompositions, was pioneered by Andersson–Blue in [77]. Their method, which extends the classical vectorfield

[32] For the analogous result in the case of the scalar wave equation, see [58]. See also [59] for a stronger quantitative version which was used in [60].

[33] The first realistic boundedness result for solutions of the scalar wave equation in Schwarzschild appears in [64] based on a clever use of the energy method which takes into account the degeneracy of $\mathbf{T}$ at the horizon.

[34] These replace the scaling and inverted time translation vectorfields used in [29] or their corresponding deformations used in [20]. A recent improvement of the method allowing one to derive higher order decay can be found in [68].

[35] In his Princeton PhD thesis Stogin establishes a Morawetz estimate even for the full subextremal case $|a| < m$.

[36] See also [75] for the first proof of boundedness of solutions, based on mode decompositions.

method to include second order operators (in this case the Carter operator, see Section 2.1), has the usual advantages of the classical vectorfield method, i.e. it is robust with respect to perturbations. It is for this reasons that [3] relies on it in the proof of Theorem 4.1.

3.3. *Linear stability of Schwarzschild*

A first quantitative proof of the linear stability of Schwarzschild spacetime was established[37] by Dafermos–Holzegel–Rodnianski (DHR) in [23]. Notable in their analysis is the treatment of the Teukolsky equation in a fixed Schwarzschild background. While the Teukolsky equation is separable, and amenable to mode analysis, it is not variational and thus cannot be treated directly by energy type estimates. As mentioned earlier in Section 3.1, Chandrasekhar was able to relate the Teukolsky equation to the Regge–Wheeler (RW) equation, which is both variational and coercive (the potential V has a favorable sign). In [23] the authors rely on a physical space version of the Chandrasekhar transformation. Once decay estimates for the RW equation have been established, based on the technology developed earlier for the scalar wave equation in Schwarzschild, the authors recover the expected boundedness and decay for solutions to the original Teukolsky equation.

The remaining work in [23] is to derive similar control for the other curvature components and the linearized Ricci coefficients associated to the double null foliation. This last step requires carefully chosen gauge conditions, which the authors make within the framework of a double null foliation, initialized both on the initial hypersurface and the background Schwarzschild horizon[38]. This gauge fixing from initial data leads to sub-optimal decay estimates for some of the metric coefficients[39] and is thus inapplicable to the nonlinear case. This deficiency was fixed in the PhD thesis of E. Giorgi, in the context of the linear stability of Reissner–Nordström, see [78,79], by relying on a linearized version of the GCM construction in [80].

3.4. *Linear stability of Kerr for small angular momentum*

The first breakthrough result on the linear stability of Kerr, for $|a|/m \ll 1$, is due to Ma [81], followed by [82]. Both results are based on a generalization of the Chandrasekhar transformation to Kerr which takes the Teukolsky equations, verified by the extreme curvature components, to generalized versions of the Regge–Wheeler (gRW) equation. Relying on separation of variables and vectorfield techniques, similar to those developed for the scalar wave equation in slowly rotating Kerr, the authors derive Energy–Morawetz and r^p estimates for the solutions of the gRW equations. Note that these results were recently extended to the full subextremal range, $|a| < m$, in [83,84] and[40] [86].

The first stability results for the full linearized Einstein vacuum equations near $\mathcal{K}(a,m)$, for $|a|/m \ll 1$, appeared in [85,87]. The first paper, based on the GHP formalism[41] see [88], builds on the results of [81] while the second paper is based on an adapted version of the metric formalism and builds on the seminal work of the authors on Kerr–de Sitter [89]. Though the ultimate relevance of these papers to nonlinear stability remains open, they are both remarkable results in so far as they deal with difficulties that looked insurmountable even ten years ago.

[37] A somewhat weaker version of linear stability of Schwarzschild was subsequently proved in [24] by using the original, direct, Regge–Wheeler, Zerilli approach combined with the vectorfield method and adapted gauge choices. See also [25] for an alternate approach of linear stability of Schwarzschild using wave coordinates.

[38] The authors use a a scalar condition for the linearized lapse along the event horizon (part of what the authors call future normalized gauge), itself initialized from initial data, see (212) and (214) in [23].

[39] See (250)–(252) and (254) in [23].

[40] This latter uses the micro-local framework of Häfner, Hintz and Vasy, see [85].

[41] An adapted spinorial version of the NP formalism.

3.5. *Nonlinear model problems*

3.5.1. *Nonlinear stability of Kerr–de Sitter*

There is another important, simplified, nonlinear model problem which has drawn attention in recent years, due mainly to the striking achievement of Hintz and Vasy [89]. This is the problem of stability of Kerr–de Sitter concerning the Einstein vacuum equation with a strictly positive cosmological constant

$$\mathbf{R}_{\alpha\beta} + \Lambda \mathbf{g}_{\alpha\beta} = 0, \quad \Lambda > 0. \tag{14}$$

In their work, which relies in part on the important mode stability result of Kodama and Ishibashi [90], Hintz and Vasy were able prove the nonlinear stability of the stationary part of Kerr–de Sitter with small angular momentum, the first nonlinear stability result of any nontrivial stationary solutions for the Einstein equations[42]. It is important to note that, despite the fact that, formally, the Einstein vacuum Equation (3) is the limit[43] of (14) as $\Lambda \to 0$, the global behavior of the corresponding solutions is radically different[44].

The main simplification in the case of stationary solutions of (14) is that the expected decay rates of perturbations near Kerr–de Sitter is exponential, while in the case $\Lambda = 0$ the decay is lower degree polynomial[45], with various components of tensorial quantities decaying at different rates, and the slowest decaying rate[46] being no better than t^{-1}. The Hintz–Vasy result was recently revisited in the work of Fang [92,93] using composite methods which bridge the gap between the spectral methods of [89] and the more conventional vectorfield methods used in the field.

3.5.2. *Nonlinear stability of Schwarzschild*

The first nonlinear stability result of the Schwarzschild space was established in [80]. In its simplest version, the result states the following.

Theorem 3.1 (Klainerman–Szeftel [80]). *The future globally hyperbolic development of an* axially symmetric, polarized, *asymptotically flat initial data set, sufficiently close (in a specified topology) to a Schwarzschild initial data set of mass $m_0 > 0$, has a complete future null infinity $\mathscr{I}^+$ and converges in its causal past $\mathscr{J}^{-1}(\mathscr{I}^+)$ to another nearby Schwarzschild solution of mass m_f close to m_0, see Figure 2 below.*

The restriction to axial polarized perturbations is the simplest assumption which insures that the final state is itself Schwarzschild and thus avoids the additional complications of the Kerr stability problem (see the introduction in [80] for a full discussion of the result).

The proof is based on an adaptation of GCM admissible spacetimes similar to that discussed below in Section 4. There are however several important simplifications to be noted:

- The assumption of polarization makes the constructions of the GCM spheres S_* and spacelike hypersurface Σ_* significantly simpler, see Chapter 9 in [80], by comparison to the general case treated in [4–6].

[42] This is also the first general nonlinear stability result in GR establishing asymptotic stability towards a family of solutions, i.e. full quantitative convergence to a final state close, but different from the initial one.

[43] To pass to the limit requires one to understand all global in time solutions of (14) with $\Lambda = 1$, not only those which are small perturbations of Kerr–de Sitter, treated by [89].

[44] Major differences between formally close equations occur in many other contexts. For example, the incompressible Euler equations are formally the limit of the Navier–Stokes equations as the viscosity tends to zero. Yet, at fixed viscosity, the global properties of the Navier–Stokes equations are radically different from that of the Euler equations.

[45] While there is exponential decay in the stationary part treated in [89], note that lower degree polynomial decay is expected in connection to the stability of the complementary causal region (called cosmological or expanding) of the full Kerr–de Sitter space, see e.g. [91].

[46] Responsible for carrying gravitational waves at large distances so that they are detectable.

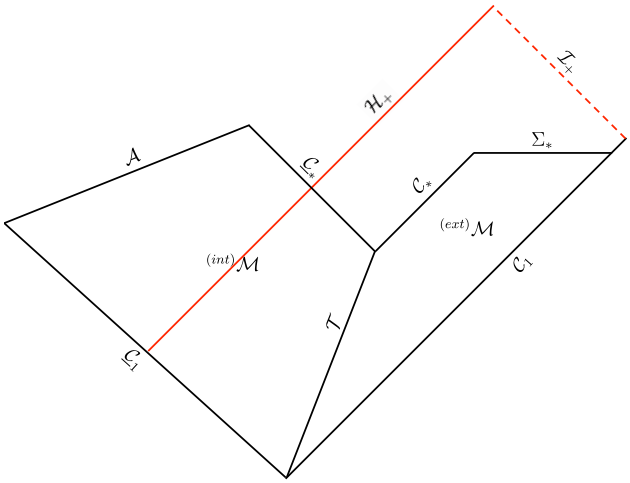


Figure 2. The GCM admissible space-time $\mathscr{M}$. By comparison to Figure 4, $\mathscr{M}$ does not have $^{(\text{top})}\mathscr{M}$, the past boundaries $\mathscr{C}_0 \cup \underline{\mathscr{C}}_0$ and future boundary $\underline{\mathscr{C}} \cup \underline{\mathscr{C}}_*$ are null and the horizontal structures (induced by geodesic foliations) are integrable. As in Theorem 4.1, the crucial GCM sphere S_* is defined and constructed with no reference to the initial data.

- The corresponding GCM admissible spacetime has only two components $\mathscr{M} = {}^{(\text{ext})}\mathscr{M} \cup {}^{(\text{int})}\mathscr{M}$ and the null horizontal structures, defined on each component, are integrable.
- As in the case of the scalar wave equation on Schwarzschild space the main spin-2 Teukolsky wave equations can be treated (via the passage to the Regge–Wheeler equation) by a vectorfield approach. This is no longer true in Kerr and even less so in perturbations of Kerr.

Recently Dafermos–Holzegel–Rodnianski–Taylor [94] have extended[47] the result of [80] by properly preparing a co-dimension 3 subset of the initial data such that the final state is still Schwarzschild. Like in [80], the starting point of [94] is to anchor the entire construction on a far away[48] GCM type sphere S_*, in the sense of [4,5], with no direct reference to the initial data. It also uses the same definition of the angular momentum as in (7.19) of [5]. Finally, the spacetime in [94] is separated in an exterior region ${}^{(\text{ext})}\mathscr{M}$ and an interior region ${}^{(\text{int})}\mathscr{M}$, with the ingoing foliation of ${}^{(\text{int})}\mathscr{M}$, initialized based on the information induced by ${}^{(\text{ext})}\mathscr{M}$, as in [80]. Note, however, that [94] does not use the geodesic foliation of [80], but instead both ${}^{(\text{int})}\mathscr{M}$ and ${}^{(\text{ext})}\mathscr{M}$ are foliated by double null foliations, and thus, the process of estimating the gauge dependent variables is somewhat different. The proof also makes use of the same definition of the angular momentum as introduced in [5] and used in [80].

4. Kerr stability for small angular momentum

Theorem 4.1. *The future globally hyperbolic development of a general, asymptotically flat, initial data set, sufficiently close (in a suitable topology) to a $\mathscr{K}(a_0, m_0)$ initial data set, for sufficiently small $|a_0|/m_0$, has a complete future null infinity $\mathscr{I}^+$ and converges in its causal past $\mathscr{J}^{-1}(\mathscr{I}^+)$*

[47] The novelty of [94], compared to [80], is the well preparation of the initial data, based on an additional three dimensional modulation. Note however that [94] requires substantially stronger asymptotic conditions for the initial data compared to [80].

[48] That is $r \gg u$, similar to the dominant in r condition (3.3.4) of [80].

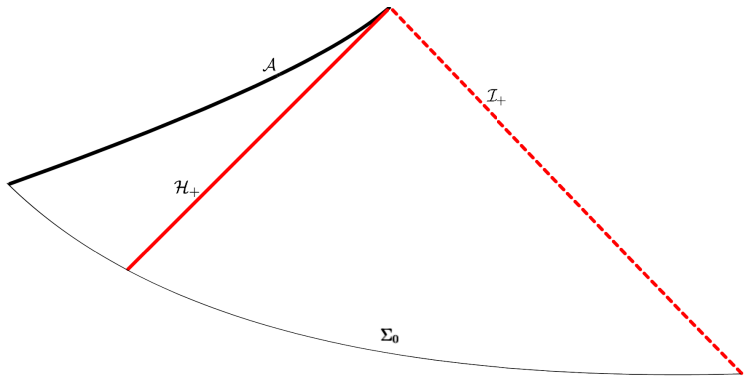


Figure 3. The Penrose diagram of the final space-time in Theorem 4.1 with initial hypersurface Σ_0, future space-like boundary $\mathscr{A}$, and $\mathscr{I}^+$ the complete future null infinity. The null hypersurface $\mathscr{H}_+$ is the future event horizon of the final Kerr. This was recently shown to be a smooth hypersurface in [95].

to another nearby Kerr spacetime $\mathscr{K}(a_f, m_f)$ with parameters (a_f, m_f) close to the initial ones (a_0, m_0), see Figure 3 below.

The precise version of the result, all the main features of the architecture of its proof, as well as detailed proofs for most of the main steps are to be found in [2]. The full proof relies also on [3] as well as [4,5] on GCM spheres, and the extension [6] to GCM hypersurfaces. In what follows I describe some of the main features of the proof.

4.1. *The bootstrap region*

The proof of Theorem 4.1 is centered around a limiting argument for a family of carefully constructed finite *generally covariant modulated (GCM) admissible spacetimes* $\mathscr{M} = {}^{(\mathrm{ext})}\mathscr{M} \cup {}^{(\mathrm{top})}\mathscr{M} \cup {}^{(\mathrm{int})}\mathscr{M}$. As can be seen in Figure 4 below, the future boundary of the spacetime is given by $\mathscr{A} \cup {}^{(\mathrm{top})}\Sigma \cup \Sigma_*$ where Σ_* is a spacelike, generally covariant modulated (GCM) hypersurface, that is a hypersurface verifying a set of crucial, well-specified, geometric conditions, essential to the proof of convergence to a final state.

The capstone as well as the most original part of the entire construction is the sphere S_*, the future boundary of Σ_*, which verifies a set of rigid, extrinsic and intrinsic, conditions. Once Σ_* is specified the whole GCM admissible spacetime $\mathscr{M}$ is determined by a more conventional construction, based on geometric transport type equations. More precisely ${}^{(\mathrm{ext})}\mathscr{M}$ can be determined from Σ_* by a specified outgoing foliation terminating in the timelike boundary $\mathscr{T}$, ${}^{(\mathrm{int})}\mathscr{M}$ is determined from $\mathscr{T}$ by a specified incoming one, and ${}^{(\mathrm{top})}\mathscr{M}$ is a complement of ${}^{(\mathrm{ext})}\mathscr{M} \cup {}^{(\mathrm{int})}\mathscr{M}$ which makes $\mathscr{M}$ a causal domain[49]. The past boundary $\mathscr{B}_1 \cup \underline{\mathscr{B}}_1$ of $\mathscr{M}$, which is itself to be constructed, is included in the initial layer $\mathscr{L}_0$ in which the spacetime is assumed to be known, i.e. a small perturbation of a Kerr solution. The passage from the initial data specified on Σ_0 to the initial layer spacetime $\mathscr{L}_0$ is justified by Shen in [96] by arguments similar to those of [34,35], based on the mathematical methods and techniques introduced in [20].

Each of the spacetime regions ${}^{(\mathrm{ext})}\mathscr{M}, {}^{(\mathrm{int})}\mathscr{M}, {}^{(\mathrm{top})}\mathscr{M}$ come equipped with specific geometric structure including specific choices of null frames and functions such as $r, u, \underline{u}$. These are first defined on Σ_* and then transported to ${}^{(\mathrm{ext})}\mathscr{M}, {}^{(\mathrm{int})}\mathscr{M}, {}^{(\mathrm{top})}\mathscr{M}$.

[49] This is required because of the fact that, in the [80] construction, the future boundary of ${}^{(\mathrm{ext})}\mathscr{M} \cup {}^{(\mathrm{int})}\mathscr{M}$ is not causal. By contrast, in [80], $\mathscr{M} = {}^{(\mathrm{ext})}\mathscr{M} \cup {}^{(\mathrm{int})}\mathscr{M}$.

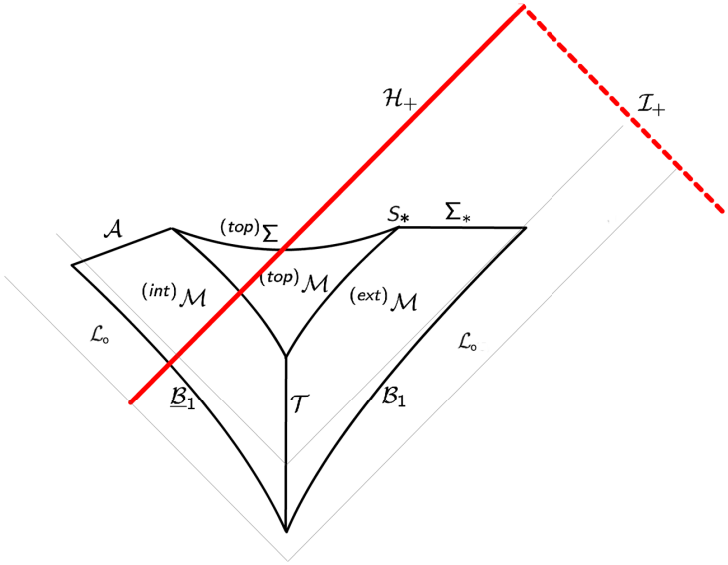


Figure 4. The Penrose diagram of a finite GCM admissible space-time $\mathscr{M} = {}^{(\mathrm{ext})}\mathscr{M} \cup {}^{(\mathrm{top})}\mathscr{M} \cup {}^{(\mathrm{int})}\mathscr{M}$. The spacetime is prescribed in the initial layer $\mathscr{L}_0$ and has $\mathscr{A} \cup {}^{(\mathrm{top})}\Sigma \cup \Sigma_*$ as future boundary, with Σ_* a spacelike "generally covariant modulated (GCM)" hypersurface. Its past boundary, $\mathscr{B}_1 \cup \underline{\mathscr{B}}_1$, is itself part of the construction. ${}^{(\mathrm{ext})}\mathscr{M}$ is initialized by the GCM hypersurface Σ_* while ${}^{(\mathrm{int})}\mathscr{M}$ is initialized on $\mathscr{T}$ by the foliation induced by ${}^{(\mathrm{ext})}\mathscr{M}$. The main inovation is the GCM sphere S_*, defined and constructed with no reference to the initial data prescribed in the initial data layer $\mathscr{L}_0$.

4.2. *Main intermediary results*

The proof of Theorem 4.1 is divided in nine separate steps, Theorems M0–M8. These steps are briefly described below, see Section 3.7 in [2] for the precise statements:

(1) *Theorem M0 (Control of the initial data in the bootstrap gauge).* The smallness of the initial perturbation is given in the frame of the initial data layer $\mathscr{L}_0$. Theorem M0 transfers this control to the bootstrap gauge in the initial data layer.

(2) *Theorems M1–M2 (Decay estimates for α (Theorem M1) and $\underline{\alpha}$ (Theorem M2)).* This is achieved using Teukolsky equations and a Chandrasekhar type transform in perturbations of Kerr.

(3) *Theorems M3–M5 (Decay estimates for all curvature, connection and metric components).* This is done making use of the GCM conditions on Σ_* as well as the control of α and $\underline{\alpha}$ established in Theorems M1 and M2. The proof proceeds in the following order:
   - Theorem M3 provides the crucial decay estimates on Σ_*,
   - Theorem M4 provides the decay estimates on ${}^{(\mathrm{ext})}\mathscr{M}$,
   - Theorem M5 provides the decay estimates on ${}^{(\mathrm{int})}\mathscr{M}$ and ${}^{(\mathrm{top})}\mathscr{M}$.

(4) *Theorems M6 (Existence of a bootstrap spacetime).* This theorem shows that there exists a GCM admissible spacetime satisfying the bootstrap assumptions, hence initializing the bootstrap procedure.

(5) *Theorems M7 (Extension of the bootstrap region).* This theorem shows the existence of a slightly larger GCM admissible spacetime satisfying estimates improving the bootstrap assumptions on decay.

(6) *Theorem M8 (Control of the top derivatives estimates).* This is based on an induction argument relative to the number of derivatives, energy–Morawetz estimates and the Maxwell like character of the Bianchi identities.

The paper [2] provides the proof of Theorem M0, Theorems M3 to M7, and half of Theorem M8 (on the control of Ricci coefficients and metric components). The proof of Theorems M1 and M2, and of the other half of Theorem M8 (on the control of curvature components), based on nonlinear wave equations techniques, are provided in [3]. The construction of GCM spheres in [4,5], and of GCM hypersurfaces in [6] are used in the proof of Theorems M6 and M7 to construct respectively the terminal GCM sphere S_* and the last slice hypersurface Σ_* from S_*.

4.3. *Main new ideas of the proof*

Here is a short description of the main new ideas in the proof of Theorem 4.1 and how they compare with ideas used in other nonlinear results.

4.3.1. *GCM admissible spacetimes*

- As mentioned already the crucial concept in the proof of Theorem 4.1 is that of a GCM admissible spacetime, whose construction is anchored by the GCM sphere S_* in Figure 4. GCM spheres[50], are codimension 2 compact surfaces, unrelated to the initial conditions, on which specific geometric quantities take Schwarzschildian values (made possible by taking into account the full general covariance of the Einstein vacuum equations). In addition to these extrinsic conditions the sphere S_* is endowed with a choice of "effective[51] isothermal coordinates", (θ,φ) verifying the following properties:
  - The metric on S_* takes the form $g = e^{2\phi} r^2((d\theta)^2 + \sin^2\theta (d\varphi)^2)$.
  - The integrals on S_* of the $\ell = 1$ modes[52] $J^{(0)} := \cos\theta$, $J^{(-)} := \sin\theta\sin\varphi$ and $J^{(+)} := \sin\theta\cos\varphi$ vanish identically.
- Given the GCM sphere S_* and the effective isothermal coordinates (θ,φ) on it, the GCM procedure allows us, in particular, to define the mass m, the angular momentum a and a virtual axis of rotation which converge, in the limit, to the final parameters a_f, m_f and the axis of rotation of the final Kerr[53]. I refer the reader to Section 7.2 in [5] for the intrinsic definition of a and of the virtual axis of symmetry on a GCM sphere.
- The boundary Σ_*, called a GCM hypersurface, is initialized at S_* and verifies additional conditions. In the polarized setting the first such construction appears [80]. The general case is treated in [6].
- The concepts of GCM spheres has appeared first in [80] in the context of polarized symmetry. The construction of GCM spheres, without any symmetries, in realistic perturbations of Kerr, is treated in [4,5][54].
- The main novelty of the GCM approach is that it relies on gauge conditions initialized at a far away co-dimension 2 sphere S_*, with no direct reference to the initial conditions. Previously known geometric constructions, such as in [20,34,101], were based on codimension-1 foliations constructed on spacelike or null hypersurfaces and initialized

[50] See the discussion in the introductions to [4,5].

[51] This is meant to insure the rigidity of the uniformization map, see [5].

[52] This is a natural generalization of $\ell = 1$ spherical harmonics.

[53] Previous definitions of the angular momentum in General Relativity were given in [97–99], see also [100] for a comprehensive discussion of the subject.

[54] See also chapter 16 of [94] in the particular case of perturbations of Schwarzschild, where the same concept appears instead under the name "teleological".

on the initial hypersurface[55]. Gauge conditions initialized from the future with no direct reference to the initial conditions, which was initiated in [80], have since been used in other works, see [41,78,79,94].

- The GCM construction introduces the following new important conceptual difficulty. The foliation on Σ_*, induced from the far away sphere S_*, needs to be connected, somehow, to the initial conditions (i.e. the initial layer $\mathscr{L}_0$ in Figure 4). This is achieved in both [80] and [2] by transporting[56] the sphere S_* to a sphere S_1 in the the initial layer and compare it, using the rigidity properties of the GCM conditions, to a sphere of the initial data layer. This induces a new foliation of the initial layer which differs substantially from the original one, due to a shift of the center of mass frame of the final black hole, known in the physics literature as a gravitational wave recoil[57].

4.3.2. *Non integrability of the horizontal structure*

As mentioned in Section 2.1, the canonical horizontal structure induced by the principal null directions (e_3, e_4) in (4) of Kerr are non integrable. The lack of integrability is dealt with by the Newman–Penrose (NP) formalism by general null frames (e_3, e_4, e_1, e_2), with e_1, e_2 a specified basis[58] of the horizontal structure induced by the null pair (e_3, e_4). It thus reduces all calculations to equations involving the Christoffel symbols of the frame, as scalar quantities. This un-geometric feature of the formalism makes it difficult to use it in the nonlinear setting of the Kerr stability problem. Indeed complex calculations depend on higher derivatives of all connection coefficients of the NP frame rather than only those which are geometrically significant. This seriously affects and complicates the structure of non-linear corrections and makes it difficult to avoid artificial gauge type singularities[59]. This difficulty is avoided in [20] by working with a tensorial approach adapted to S-foliations, i.e. $\{e_3, e_4\}^\perp$ coincides, at every point, with the tangent space to S.

The paper [3] extends, with minimal changes, the tensorial approach introduced in [20], to general non-integrable foliations. The idea is simply to define Ricci coefficients $\chi, \underline{\chi}, \eta, \underline{\eta}, \zeta, \xi, \underline{\xi}, \omega, \underline{\omega}$ exactly as in [20], relative to an arbitrary basis of vectors $(e_a)_{a=1,2}$ of $\mathscr{H} := \{e_3, e_4\}^\perp$. In particular, the null fundamental forms χ and $\underline{\chi}$, are given by

$$\underline{\chi}_{ab} = \mathbf{g}(\mathbf{D}_a e_3, e_b), \quad \chi_{ab} = \mathbf{g}(\mathbf{D}_a e_4, e_b).$$

Due to the lack of integrability of $\mathscr{H}$, the null fundamental forms χ and $\underline{\chi}$ are no longer symmetric. They can be both decomposed as follows

$$\chi_{ab} = \tfrac{1}{2}\mathrm{tr}\,\chi\delta_{ab} + \tfrac{1}{2}\in_{ab}{}^{(a)}\mathrm{tr}\chi + \widehat{\chi}_{ab}, \quad \underline{\chi}_{ab} = \tfrac{1}{2}\mathrm{tr}\,\underline{\chi}\delta_{ab} + \tfrac{1}{2}\in_{ab}{}^{(a)}\mathrm{tr}\underline{\chi} + \widehat{\underline{\chi}}_{ab},$$

where the new scalars ${}^{(a)}\mathrm{tr}\chi$, ${}^{(a)}\mathrm{tr}\underline{\chi}$ measure the lack of integrability of the horizontal structure. The null curvature components are also defined as in [20],

$$\alpha_{ab} = \mathbf{R}_{a4b4}, \quad \beta_a = \tfrac{1}{2}\mathbf{R}_{a434}, \quad \underline{\beta}_a = \tfrac{1}{2}\mathbf{R}_{a334}, \quad \underline{\alpha}_{ab} = \mathbf{R}_{a3b3}, \quad \rho = \tfrac{1}{4}\mathbf{R}_{3434}, \quad {}^*\rho = \tfrac{1}{4}{}^*\mathbf{R}_{3434}.$$

The null structure and null Bianchi equations can then be derived as in the integrable case, see chapter 7 in [20]. The only new features are the presence of the scalars ${}^{(a)}\mathrm{tr}\chi$, ${}^{(a)}\mathrm{tr}\underline{\chi}$ in the

[55] The first such construction appears in the proof of the nonlinear stability of the Minkowski space [20] where the "inverse lapse foliation" was constructed on the "last slice", initialized at spacelike infinity i^0. Similar constructions, where the last slice is null rather than spacelike, appear in [34,101].

[56] That is, one transport sthe $\ell = 1$ modes of some quantities from S_* to S_1, see Section 8.3.1 in [2].

[57] See section 8.3 in [2] for the details.

[58] Or rather the complexified vectors $m = e_1 + \mathrm{i}e_2$ and $\overline{m} = e_1 - \mathrm{i}e_2$.

[59] There are no smooth, global choices of a basis (e_1, e_2). The choice (5) in Kerr, for example, is singular at $\theta = 0, \pi$.

equations. Finally note that the equations acquire additional simplicity if we pass to complex notations[60],

$$\begin{aligned} &A := \alpha + \mathrm{i}^*\alpha, \quad B := \beta + \mathrm{i}^*\beta, \quad P := \rho + \mathrm{i}^*\rho, \quad \underline{B} := \underline{\beta} + \mathrm{i}^*\underline{\beta}, \quad \underline{A} := \underline{\alpha} + \mathrm{i}^*\underline{\alpha}, \\ &X := \chi + \mathrm{i}^*\chi, \quad \underline{X} := \underline{\chi} + \mathrm{i}^*\underline{\chi}, \quad H := \eta + \mathrm{i}^*\eta, \quad \underline{H} := \underline{\eta} + \mathrm{i}^*\underline{\eta}, \quad Z := \zeta + \mathrm{i}^*\zeta. \end{aligned} \tag{15}$$

4.3.3. *Frame transformations and choice of frames*

Given an arbitrary perturbation of Kerr, there is no a-priori reason to prefer an horizontal structure to any other one obtained from the first by another perturbation of the same size. It is thus essential that one considers all possible frame transformations from one horizontal structure $(e_4, e_3, \mathscr{H})$ to another one $(e_4', e_3', \mathscr{H}')$ together with the transformation formulas $\Gamma \to \Gamma'$, $R \to R'$ they generate. The most general transformation formulas between two null frames is given in Lemma 3.1 of [4]. It depends on two horizontal 1-forms $f, \underline{f}$ and a real scalar function λ and is given by

$$\begin{aligned} e_4' &= \lambda\Big(e_4 + f^b e_b + \tfrac{1}{4}|f|^2 e_3\Big), \\ e_a' &= \Big(\delta_a^b + \tfrac{1}{2}\underline{f}_a f^b\Big) e_b + \tfrac{1}{2}\underline{f}_a e_4 + \Big(\tfrac{1}{2} f_a + \tfrac{1}{8}|f|^2 \underline{f}_a\Big) e_3, \\ e_3' &= \lambda^{-1}\Big(\Big(1 + \tfrac{1}{2} f\cdot\underline{f} + \tfrac{1}{16}|f|^2|\underline{f}|^2\Big) e_3 + \Big(\underline{f}^b + \tfrac{1}{4}|\underline{f}|^2 f^b\Big) e_b + \tfrac{1}{4}|\underline{f}|^2 e_4\Big). \end{aligned} \tag{16}$$

The very important transformation formulas $\Gamma \to \Gamma'$, $R \to R'$ are given in Proposition 3.3 of [4].

The case of $\mathscr{K}(a,m)$, $a \neq 0$ presents an interesting new feature which can be described as follows:

- To capture the simplicity induced by the principle null directions in Kerr it is natural to work with non-integrable frames and define all main quantities relative to adapted frames for which all quantities which vanish in Kerr are of the size of the perturbation.
- A crucial aspect of all important results in GR, based on integrable S- foliations, is that one can rely on elliptic Hodge theory on each 2-surface S. This is no longer possible in context where the main quantities and the basic equations they verify are defined relative to non integrable frames. In [2,3] this problem is dealt with by passing back and forth, whenever needed, from the main non-integrable frame to a well chosen adapted integrable frame, according to the transformation formulas mentioned above.

4.3.4. *Renormalization procedure and the canonical complex 1-form* $\mathfrak{J}$

The main complex quantities introduced in (15) take a particularly simple form in the principal null frame (4) of Kerr:

$$\begin{aligned} &A = \underline{A} = B = \underline{B} = 0, \quad P = -\frac{2m}{q^3}, \\ &\widehat{X} = \widehat{\underline{X}} = 0, \quad \mathrm{tr}X = \frac{2}{q}\frac{\Delta}{|q|^2}, \quad \mathrm{tr}\underline{X} = -\frac{2}{\overline{q}}, \\ &Z = \frac{aq}{|q|^2}\mathfrak{J}, \quad H = \frac{aq}{|q|^2}\mathfrak{J}, \quad \underline{H} = -\frac{a\overline{q}}{|q|^2}\mathfrak{J}, \end{aligned} \tag{17}$$

where $q = r + \mathrm{i}a\cos\theta$, and where the regular[61] complex 1-form $\mathfrak{J}$ is given by

$$\mathfrak{J}_1 = \frac{\mathrm{i}\sin\theta}{|q|}, \quad \mathfrak{J}_2 = \frac{\sin\theta}{|q|}, \tag{18}$$

see Sections 2.4.2 and 2.4.3 in [2]. In particular, the following holds for the complexified horizontal tensors of (15) in the principal null frame (4) of Kerr:

[60] The dual here is taken with respect to the antisymmetric horizontal 2-tensor ϵ_{ab}.

[61] Note that $\mathfrak{J}$ is regular including at $\theta = 0, \pi$.

- the complex scalars P, $\mathrm{tr}X$ and $\mathrm{tr}\underline{X}$ are functions of r and $\cos\theta$,
- the non vanishing complex 1-forms H, $\underline{H}$ and Z consist of functions of r and $\cos\theta$ multiplied by $\mathfrak{J}$,
- the traceless symmetric complex 2-tensors A, $\underline{A}$, $\widehat{X}$ and $\widehat{\underline{X}}$ vanish identically.

Based on that observation, for a given horizontal structure perturbing the one of Kerr, one can define a renormalization procedure by which, once we have[62] suitable constants (a,m), suitable scalar functions (r,θ), and a suitable complex 1-form $\mathfrak{J}$, and after subtracting the corresponding values in Kerr computed from $(a,m,r,\theta,\mathfrak{J})$ for all the Ricci and curvature coefficients, we obtain quantities which are first order in the perturbation.

More precisely, once (a,m), (r,θ) and $\mathfrak{J}$ have been chosen, one renormalizes the quantities in (15) that do not vanish in Kerr as follows[63]:

$$\begin{aligned} &\check{P} := P + \frac{2m}{q^3}, \quad \mathrm{tr}\check{X} := \mathrm{tr}X - \frac{2}{q}\frac{\Delta}{|q|^2}, \quad \mathrm{tr}\check{\underline{X}} := \mathrm{tr}\underline{X} + \frac{2}{\overline{q}}, \\ &\check{Z} := Z - \frac{aq}{|q|^2}\mathfrak{J}, \quad \check{H} := H - \frac{aq}{|q|^2}\mathfrak{J}, \quad \check{\underline{H}} := \underline{H} + \frac{a\overline{q}}{|q|^2}\mathfrak{J}. \end{aligned} \tag{19}$$

4.3.5. *Principal geodesic and principal temporal structures*

In addition to the GCM gauge conditions on Σ_*, one needs to construct a gauge on $\mathscr{M}$ which relates the non integrable horizontal structure to the scalars (r,θ) and the complex 1-form $\mathfrak{J}$. Two such gauges were introduced in [2]:

- Principal Geodesic (PG) structure, which is a generalization of the geodesic foliation to non-integrable horizontal structures,
- Principal Temporal (PT) structure, which favors transport equations along a null direction.

The PG structure[64] is well suited for decay estimates, but fails to be well posed. Indeed, due to the lack of integrability of the horizontal structure, we cannot control the null structure equations[65] without a loss of derivative. The PT structure, on the other hand, is designed so that the loss of derivatives in the null structure equations, in the incoming or outgoing direction, is completely avoided. Note however that the PT structure is not well suited to the derivation of decay estimates on ${}^{(\mathrm{ext})}\mathscr{M}$ where r can take arbitrary large values. In [2] the authors work with both gauge conditions, depending on the goal they want to achieve, and rely on the transformation formulas (16) to pass from one to the other.

In the outgoing normalization both the outgoing PG and PT structures consist of a choice $(e_3, e_4, \mathscr{H})$, with e_4 null geodesic, together with a scalar functions r,θ and a complex 1-form $\mathfrak{J}$ such that $e_4(r) = 1$, $e_4(\theta) = 0$, $\nabla_4(q\mathfrak{J}) = 0$ where $q = r + \mathrm{i}a\cos\theta$. In addition:

(1) In a PG structure the gradient of r, given by $N = \mathbf{g}^{\alpha\beta}\partial_\beta r\partial_\alpha$, is perpendicular to $\mathscr{H}$,
(2) In a PT structure $\underline{H} = -a\overline{q}/|q|^2\mathfrak{J}$, i.e. $\check{\underline{H}} = 0$ in view of (19).

A similar definition of incoming PG and PT structures is obtained by interchanging the roles of e_3, e_4. Note that both structures still need to be initialized. The outgoing PG and PT structures of ${}^{(\mathrm{ext})}\mathscr{M}$ are both initialized on Σ_* from the GCM frame of Σ_*, while the ingoing PT structures of

[62] The constants m and a are computed on the GCM sphere S_*, see Section 4.3.1. r, θ and $\mathfrak{J}$ are chosen on S_*, transported to Σ_* and then to $\mathscr{M}$. The horizontal structure is also defined first on Σ_* and then transported to $\mathscr{M}$.

[63] The renormalization is written here in the case of a null pair (e_3, e_4) with an ingoing normalization.

[64] Note that, in the integrable context of [80], the PG structure coincides with the standard (integrable) geodesic foliation used there. Thus the PG structure, defined in [2], is a suitable generalization to the non-integrable case of perturbations of Kerr.

[65] In integrable situation, like in the case of S-foliations, the Hodge systems on the leaves of the S-foliation allows us to avoid the loss.

${}^{(\mathrm{int})}\mathscr{M}$ and ${}^{(\mathrm{top})}\mathscr{M}$ are initialized on the the timelike hypersurface $\mathscr{T}$, see Figure 4, using the data induced by the outgoing structures.

4.3.6. *Control of the extreme curvature components $A, \underline{A}$*

It was already observed by Teukolsky that, in linear theory, the extreme components of the curvature are both gauge invariant and verify decoupled wave equations[66]. In the nonlinear context this translates to the statement that the horizontal 2-tensors $A, \underline{A}$, defined relative to an $O(\epsilon)$ perturbation of the principal frame of Kerr, are $O(\epsilon^2)$-invariant, relative to $O(\epsilon)$ frame transformations[67], and verify tensorial wave equations of the form

$$\dot{\Box}_2 A + L[A] = \mathrm{Err}(\check{\Gamma}, \check{R}), \quad \dot{\Box}_2 \underline{A} + \underline{L}[\underline{A}] = \underline{\mathrm{Err}}(\check{\Gamma}, \check{R}). \tag{20}$$

Here $\dot{\Box}_2$ denotes the wave operator on horizontal symmetric traceless 2-tensors, L and $\underline{L}$ are linear first order operators and $\check{\Gamma}, \check{R}$ denote the linearized Ricci and curvature coefficients. The error terms $\mathrm{Err}(\check{\Gamma}, \check{R})$, $\underline{\mathrm{Err}}(\check{\Gamma}, \check{R})$ are nonlinear expressions in $\check{\Gamma}, \check{R}$.

In linear theory, i.e. when $\mathbf{g}$ is the Kerr metric and the error terms are not present, these equations have been treated by [23] in Schwarzschild[68] and by [81,82] in slowly rotating[69] Kerr, i.e. $|a|/m \ll 1$. More precisely both results derive realistic decay estimates for $A, \underline{A}$. The methods are however not robust. Indeed, a crucial ingredient in the proof, the Energy–Morawetz estimates, is based on separation of variables. The control of A and $\underline{A}$ in perturbations of Kerr in [3] contains the following new features:

- *Derivation of the gRW equation.* The derivation of the generalized Regge–Wheeler equations in Kerr, in [81,82], is done starting with the complex, scalar, Teukolsky equations, derived via the NP, or GHP formalism, by applying a Chandrasekhar type transformation. Part I of [3] extends their derivation, using the non-integrable horizontal formalism, to perturbations of Kerr. By contrast with [3,81,82] derives gRW equations for the horizontal symmetric traceless 2-tensors[70] $\mathfrak{q}, \underline{\mathfrak{q}}$, rather than for complex scalars. The main difficulty here is to make sure that the non-linear error terms verify a favorable structure.
- *Nonlinear error terms.* The control of the nonlinear terms and their associated null structure was already understood in perturbations of Schwarzschild in [80] and is extended to perturbations of Kerr in [3].
- *Energy–Morawetz.* To derive energy–morawetz estimates for $A, \underline{A}$, [3] (Part II) vastly extends the pioneering idea of Andersson and Blue [77], based on commutations with $\mathbf{T}, \mathbf{Z}$ and the second order Carter operator $\mathscr{C}$, developed in the context of the scalar wave equation in slowly rotating Kerr, to treat the tensorial Teukolsky and gRW equations in perturbations of Kerr.

4.3.7. *Comment on the full sub-extremal range*

Though the full sub-extremal range $|a| < m$ remains open, a large part of the proof of stability does not require the smallness of $|a|/m$. This is the case for [2,4–6]. In fact the smallness assumption is only needed in [3], mostly in the derivation of the main energy–morawetz estimates in parts II and III.

[66] See discussion in Section 2.6.
[67] This means that $f, \underline{f}, \lambda - 1$ are $O(\epsilon)$ in the transformation formulas (16).
[68] See discussion in Section 3.3.
[69] See discussion in Section 3.4.
[70] Derived from $A, \underline{A}$, see Definition 5.2.2 and 5.3.3 in [3].

5. Stability of the Kerr–Newman family

As stated in the introduction I make only passing remarks about the stability the Kerr–Neman family $\mathcal{KN}(a,m,e)$. To start with, as discussed in [102], unlike in the case of Kerr, mode stability fails for Kerr–Neman spacetimes due to lack of separability and the coupling of the gravitational radiation with the electromagnetic one. So far

As mentioned earlier (see Section 3.3 the first results on the linear stability of the Reissner–Nordström family, i.e. special case of Kerr–Newman with $a = 0$, are due to Giorgi in [78,79], based on an approach similar to that used in [80]. The main novelty of those works were to identify (1)) the correct gauge invariant quantities (in fact a couple of tensors), (2) the correct Teukolsky-type coupled systems they verify, (3) the correct version of the Chandrasekhar transformation and the correct generalization of the Regge–Wheeler type system into which the Teukolsky type system is transformed (4) Derivation of Morawetz-Energy type estimates for this system. The same steps were than carried out in the more difficult case of Kerr–Newman [103–105] in which linear stability is established for $|a|,|Q| \ll M$. Recently in [106] the authors derive boundedness and decay for the Giorgi–Teukolsky system, in axial symmetry, in the range $|a| \ll M, |Q| < M$. The linear stability of the Kerr–Newman family for $|a|,|Q| \ll M$. was also established by He [107] in the metric framework using the powerful micro-local techniques of Hintz and Vasy.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

Supported by NSF grant 220103 the authors also likes to thank IHES for hosting him during his sabbatical for Princeton, February 1–May 31.

References

[1] Y. Fourès-Bruhat, “Théorème d’existence pour certains systémes d’equations aux dérivées partielles non-linéaires”, *Acta Math.* **88** (1952), pp. 141–225.

[2] S. Klainerman and J. Szeftel, “Kerr stability for small angular momentum”, *Pure Appl. Math. Quat. (PAMQ)* **19** (2023), no. 3, pp. 791–1678.

[3] E. Giorgi, S. Klainerman and J. Szeftel, “Wave equations estimates and the nonlinear stability of slowly rotating Kerr black holes”, *Pure Appl. Math. Quat. (PAMQ)* **20** (2024), no. 7, pp. 2865–3849.

[4] S. Klainerman and J. Szeftel, “Construction of GCM spheres in perturbations of Kerr”, *Ann. PDE* **8** (2022), article no. 17.

[5] S. Klainerman and J. Szeftel, “Effective results in uniformization and intrinsic GCM spheres in perturbations of Kerr”, *Ann. PDE* (2022), no. 8, article no. 18.

[6] D. Shen, “Construction of GCM hypersurfaces in perturbations of Kerr”, *Ann. PDE* **9** (2023), article no. 11.

[7] R. P. Kerr, “Gravitational field of a spinning mass as an example of algebraically special metrics”, *Phys. Rev. Lett.* **11** (1963), pp. 237–238.

[8] E. Newman and A. Janis, “Note on the Kerr spinning-particle metric”, *J. Math. Phys.* **6** (1965), no. 6, pp. 915–917.

[9] A. Ionescu and S. Klainerman, “Rigidity results in general relativity: a review”, *Surv. Differ. Geometry* **20** (2015), pp. 123–156.

[10] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society: Zürich, 2009.

[11] S. Klainerman and I. Rodnianski, “On the break-down criterion in general relativity”, *J. Am. Math. Soc.* **23** (2010), pp. 345–382.

[12] X. An and Q. Han, "Anisotropic dynamical horizons arising in gravitational collapse", preprint, 2010.12524. 51 pp.

[13] X. An, "Naked singularity censoring with anisotropic apparent horizon", *Ann. Math.*. (in press).

[14] D. Christodoulou, "On the global initial value problem and the issue of singularities", *Class. Quantum Grav.* **16** (1999), A23–A35.

[15] B. Carter, "Global structure of the Kerr family of gravitational fields", *Phys. Rev.* **174** (1968), pp. 1559–1571.

[16] S. Chandrasekhar, "An introduction to the theory of the Kerr metric and its perturbations", in *General Relativity, an Einstein Centenary Survey* (S. W. Hawking and W. Israel, eds.), Cambridge University Press: Cambridge, 1979.

[17] R. Donninger, J. Krieger, J. Szeftel and W. Wong, "Codimension one stability of the catenoid under the vanishing mean curvature flow in Minkowski space", *Duke Math. J.* **165** (2016), pp. 723–791.

[18] Y. Martel and F. Merle, "Asymptotic stability of solitons for subcritical generalized KdV equations", *Arch. Ration. Mech. Anal.* **157** (2001), pp. 219–254.

[19] F. Merle and P. Raphaël, "On universality of blow-up profile for L^2 critical nonlinear Schrödinger equation", *Invent. Math.* **156** (2004), pp. 565–672.

[20] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton University Press: Princeton, NJ, 1993.

[21] D. Christodoulou and S. Klainerman, "Asymptotic properties of linear field theories in Minkowski space", *Commun. Pure Appl. Math.* **43** (1990), pp. 137–199.

[22] S. A. Teukolsky, "Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations", *Astrophys. J.* **185** (1973), pp. 635–648.

[23] M. Dafermos, G. Holzegel and I. Rodnianski, "Linear stability of the Schwarzschild solution to gravitational perturbations", *Acta Math.* **222** (2019), pp. 1–214.

[24] P. K. Hung, J. Keller and M. T. Wang, "Linear stability of Schwarzschild spacetime: decay of metric coefficients", *J. Differ. Geom.* **116** (2020), pp. 481–541.

[25] T. W. Johnson, "The linear stability of the Schwarzschild solution to gravitational perturbations in the generalised wave gauge", *Ann. PDE* **5** (2019), no. 2, article no. 13.

[26] B. Whiting, "Mode stability of the Kerr black hole", *J. Math. Phys.* **30** (1989), pp. 1301–1305.

[27] C. Morawetz, "Decay of solutions of the exterior initial boundary value problem for the wave equation", *Commun. Pure Appl. Math.* **14** (1961), pp. 561–568.

[28] C. Morawetz, J. V. Ralston and W. A. Strauss, "Decay of solutions of the wave equation outside non-trapping obstacles", *Commun. Pure Appl. Math.* **30** (1977), pp. 447–508.

[29] S. Klainerman, "Uniform decay estimates and the Lorentz invariance of the classical wave equations", *Commun. Pure Appl. Math.* **38** (1985), pp. 321–332.

[30] S. Klainerman, "Remarks on the global Sobolev inequalities", *Commun. Pure Appl. Math.* **40** (1987), pp. 111–117.

[31] S. Klainerman, "Long time behavior of solutions to nonlinear wave equations", in *Proceedings of the International Congress of Mathematicians (Warsaw, 1983)*, PWN: Warsaw, 1984, pp. 1209–1215.

[32] D. Christodoulou, "Global solutions of nonlinear hyperbolic equations for small initial data", *Commun. Pure Appl. Math.* **39** (1986), pp. 267–282.

[33] S. Klainerman, "The Null Condition and global existence to nonlinear wave equations", *Lect. Appl. Math.* **23** (1986), pp. 293–326.

[34] S. Klainerman and F. Nicolo, *The Evolution Problem in General Relativity*, Progress in Mathematical Physics, Birkhauser: Boston, 2003, pp. +385.

[35] S. Klainerman and F. Nicolo, "Peeling properties of asymptotic solutions to the Einstein vacuum equations", *Class. Quantum Grav.* **20** (2003), pp. 3215–3257.

[36] H. Lindblad and I. Rodnianski, "Global existence in the Einstein Vacuum equations in wave co-ordinates", *Commun. Math. Phys.* **256** (2005), pp. 43–110.

[37] L. Bieri, *An extension of the stability theorem of the Minkowski space in general relativity*, PhD thesis, ETH, 2007.

[38] H. Lindblad, "On the asymptotic behavior of solutions to Einstein's vacuum equations in wave coordinates", *Commun. Math. Phys.* **353** (2017), pp. 135–184.

[39] C. Huneau, "Stability of Minkowski spacetime with a translation space-like Killing field", *Ann. PDE* **4** (2018), article no. 12.

[40] P. Hintz and A. Vasy, "Stability of Minkowski space and polyhomogeneity of the metric", *Ann. PDE* **6** (2020), article no. 2.

[41] O. Graf, "Global nonlinear stability of Minkowski space for spacelike-characteristic initial data", preprint, 2010.12434.

[42] D. Shen, "Global stability of Minkowski space with minimal decay", preprint, 2310.07483.

[43] L. Bieri and N. Zipser, *Extensions of the stability theorem of the Minkowski space in general relativity*, AMS/IP Studies in Advanced Mathematics, American Mathematical Society: Providence, RI, 2009, pp. xxiv+491.

[44] D. Fajman, J. Joudioux and J. Smulevici, "The stability of the Minkowski space for the Einstein–Vlasov system", *Anal. PDE* **14** (2021), pp. 425–531.

[45] H. Lindblad and M. Taylor, "Global stability of Minkowski space for the Einstein–Vlasov system in the harmonic gauge", *Arch. Rational Mech. Anal.* **235** (2020), pp. 517–633.

[46] L. Bigorgne, D. Fajman, J. Joudioux, J. Smulevici and M. Thaller, "Asymptotic stability of Minkowski spacetime with non-compactly massless Vlasov matter", *Arch. Rational Mech. Anal.* **242** (2021), pp. 1–147.

[47] P. G. LeFloch and Y. Ma, "The global nonlinear stability of Minkowski space for self-gravitating massive fields", *Commun. Math. Phys.* **346** (2016), pp. 603–665.

[48] A. Ionescu and B. Pausader, *The Einstein–Klein–Gordon Coupled System: Global Stability of the Minkowski Solution*, Annals of Math Studies, Princeton University Press: Princeton NJ, 2022, pp. xi+297.

[49] T. Regge and J. A. Wheeler, "Stability of a Schwarzschild singularity", *Phys. Rev.* **108** (1957), no. 2, pp. 1063–1069.

[50] C. V. Vishveshwara, "Stability of the Schwarzschild metric", *Phys. Rev. D* **1** (1970), pp. 2870–2879.

[51] F. J. Zerilli, "Effective potential for even-parity Regge–Wheeler gravitational perturbation equations", *Phys. Rev. Lett.* **24** (1970), pp. 737–738.

[52] V. Moncrief, "Gravitational perturbations of spherically symmetric systems. I. The exterior problem", *Ann. Phys.* **88** (1975), pp. 323–342.

[53] J. M. Bardeen and W. H. Press, "Radiation fields in the Schwarzschild background", *J. Math. Phys.* **14** (1973), pp. 7–19.

[54] W. Press and S. A. Teukolsky, "Perturbations of a rotating black hole. II. Dynamical stability of the Kerr metric", *Astrophys. J.* **185** (1973), pp. 649–673.

[55] S. Chandrasekhar, "On the equations governing the perturbations of the Schwarzschild black hole", *Proc. R. Soc. Lond. A Math.* **343** (1975), pp. 289–298.

[56] R. M. Wald, "Construction of solutions of gravitational, electromagnetic, or other perturbation equations from solutions of decoupled equations", *Phys. Rev. Lett.* **41** (1978), pp. 203–206.

[57] S. Aksteiner, L. Andersson, T. Bäckdahl, A. G. Shah and B. F. Whiting, "Gauge-invariant perturbations of Schwarzschild spacetime", preprint, 1611.08291.

[58] F. Finsler, N. Kamran, J. Smoller and S.-T. Yau, "Decay of solutions of the wave equation in the Kerr geometry", *Commun. Math. Phys.* **264** (2006), pp. 465–503.

[59] Y. Shlapentokh-Rothman, "Quantitative mode stability for the wave equation on the Kerr spacetime", *Ann. Henri Poincaré* **16** (2015), pp. 289–345.

[60] M. Dafermos, I. Rodnianski and Y. Shlapentokh-Rothman, "Decay for solutions of the wave equation on Kerr exterior spacetimes *iii*: The full subextremal case $|a| < m$", *Ann. Math.* **183** (2016), pp. 787–913.

[61] L. Andersson, S. Ma, C. Paganini and B. Whiting, "Mode stability on the real axis", *J. Math. Phys.* **58** (2017), article no. 072501.

[62] F. Finsler and J. Smoller, "Linear stability of the non-extreme Kerr black hole", *Adv. Theor. Math. Phys.* **21** (2017), pp. 1991–2085.

[63] R. Teixeira da Costa, "Mode stability for the Teukolsky equation on extremal and subextremal Kerr spacetimes", *Commun. Math. Phys.* **378** (2020), pp. 705–781.

[64] B. S. Kay and R. M. Wald, "Linear stability of Schwarzschild under perturbations which are non-vanishing on the bifurcation 2-sphere", *Class. Quantum Grav.* **4** (1987), pp. 893–898.

[65] P. Blue and A. Soffer, "Semilinear wave equations on the Schwarzschild manifold. I. Local decay estimates", *Adv. Differ. Equ.* **8** (2003), pp. 595–614.

[66] M. Dafermos and I. Rodnianski, "The red-shift effect and radiation decay on black hole spacetimes", *Commun. Pure Appl. Math.* **62** (2009), pp. 859–919.

[67] M. Dafermos and I. Rodnianski, "A new physical-space approach to decay for the wave equation with applications to black hole spacetimes", in *XVIth International Congress on Mathematical Physics*, World Scientific Publishing: Hackensack, NJ, 2010, pp. 421–432.

[68] Y. Angelopoulos, S. Aretakis and D. Gajic, "A vector field approach to almost-sharp decay for the wave equation on spherically symmetric, stationary spacetimes", *Ann. PDE* **4** (2018), article no. 15.

[69] P. Blue and A. Soffer, "Errata for "Global existence and scattering for the nonlinear Schrödinger equation on Schwarzschild manifolds" , "Semilinear wave equations on the Schwarzschild manifold I: Local Decay Estimates", and "The wave equation on the Schwarzschild metric II: Local Decay for the spin 2 Regge Wheeler equation"", preprint, gr-qc/0608073. 6 pages.

[70] P. Blue and J. Sterbenz, "Uniform decay of local energy and the semi-linear wave equation on Schwarzschild space", *Commun. Math. Phys.* **268** (2006), pp. 481–504.

[71] J. Marzuola, J. Metcalfe, D. Tataru and M. Tohaneanu, "Strichartz estimates on Schwarzschild black hole backgrounds", *Commun. Math. Phys.* **293** (2010), pp. 37–83.

[72] A. Ionescu and S. Klainerman, "On the global stability of the wave-map equation in Kerr spaces with small angular momentum", *Ann. PDE* **1** (2015), article no. 1.

[73] J. Stogin, *Nonlinear wave dynamics in black hole spacetimes*, PhD thesis, Princeton University: Princeton, NJ, 2017.

[74] S. Alinhac, "Energy multipliers for perturbations of the Schwarzschild metric", *Commun. Math. Phys.* **288** (2009), pp. 199–224.

[75] M. Dafermos and I. Rodnianski, "A proof of the uniform boundedness of solutions to the wave equation on slowly rotating Kerr backgrounds", *Invent. Math.* **185** (2011), pp. 467–559.

[76] D. Tataru and M. Tohaneanu, "A local energy estimate on Kerr black hole backgrounds", *Int. Math. Res. Not.* **2** (2011), pp. 248–292.

[77] L. Andersson and P. Blue, "Hidden symmetries and decay for the wave equation on the Kerr spacetime", *Ann. Math.* **182** (2015), no. 2, pp. 787–853.

[78] E. Giorgi, "The linear stability of Reissner–Nordström spacetime for small charge", *Ann. PDE* **6** (2020), article no. 8.

[79] E. Giorgi, "The linear stability of Reissner–Nordström spacetime: the full subextremal range $|Q| < M$", *Commun. Math. Phys.* **13** (2020), pp. 1313–1360.

[80] S. Klainerman and J. Szeftel, *Global Non-Linear Stability of Schwarzschild Spacetime under Polarized Perturbations*, Annals of Math Studies, Princeton University Press: Princeton NJ, 2020, pp. xviii+856.

[81] S. Ma, "Uniform energy bound and Morawetz estimate for extreme components of spin fields in the exterior of a slowly rotating Kerr black hole II: linearized gravity", *Commun. Math. Phys.* **377** (2020), pp. 2489–2551.

[82] M. Dafermos, G. Holzegel and I. Rodnianski, "Boundedness and decay for the Teukolsky equation on Kerr spacetimes I: The case $|a| \ll M$", *Ann. PDE* **5** (2019), article no. 2.

[83] Y. Shlapentokh-Rothman and R. Teixeira da Costa, "Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: frequency space analysis", preprint, 2007.07211.

[84] Y. Shlapentokh-Rothman and R. Teixeira da Costa, "Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: physical space analysis", preprint, 2302.08916.

[85] D. Häfner, P. Hintz and A. Vasy, "Linear stability of slowly rotating Kerr black holes", *Invent. Math.* **223** (2021), pp. 1227–1406.

[86] P. Millet, "Optimal decay for solutions of the Teukolsky equation on the Kerr metric for the full subextremal range $|a| < M$", preprint, 2302.06946.

[87] L. Andersson, T. Bäckdahl, P. Blue and S. Ma, "Stability for linearized gravity on the Kerr spacetime", preprint, 1903.03859.

[88] R. Geroch, A. Held and R. Penrose, "A space-time calculus based on pairs of null directions", *J. Math. Phys.* **14** (1973), pp. 874–881.

[89] P. Hintz and A. Vasy, "The global non-linear stability of the Kerr-de Sitter family of black holes", *Acta Math.* **220** (2018), pp. 1–206.

[90] H. Kodama and A. Ishibashi, "A master equation for gravitational pertur bations of maximally symmetric black holes in higher dimensions", *Progr. Theor. Phys.* **110** (2003), pp. 701–722.

[91] V. Schlue, "Decay of the Weyl curvature in expanding black hole cosmologies", *Ann. PDE* **8** (2022), article no. 9.

[92] A. J. Fang, "Linear stability of the slowly-rotating Kerr-de Sitter family", preprint, 2207.07902.

[93] A. J. Fang, "Nonlinear stability of the slowly-rotating Kerr-de Sitter family", preprint, 2112.07183.

[94] M. Dafermos, G. Holzegel, I. Rodnianski and M. Taylor, "The non-linear stability of the Schwarzschild family of black holes", preprint, 2104.0822.

[95] X. Chen and S. Klainerman, "Regularity of the future event horizon in perturbations of Kerr", preprint, 2024, 2409.05700v1. [gr-qc] 9 Sep.

[96] D. Shen, "Kerr stability in external regions", *Ann. PDE* **10** (2024), article no. 9.

[97] A. Rizzi, "Angular momentum in general relativity: a new definition", *Phys. Rev. Lett.* **81** (1998), no. 6, pp. 1150–1153.

[98] P.-N. Chen, M.-T. Wang and S.-T. Yau, "Quasilocal angular momentum and center of mass in general relativity", *Adv. Theor. Math. Phys.* **20** (2016), pp. 671–682.

[99] P.-N. Chen, M.-T. Wang, Y.-K. Wang and S.-T. Yau, "Supertranslation invariance of angular momentum", *Adv. Theor. Math. Phys.* **25** (2021), pp. 777–789.

[100] L. B. Szabados, "Quasi-local energy-momentum and angular momentum in general relativity", *Living Rev. Relativ.* **12** (2009), article no. 4.

[101] S. Klainerman, J. Luk and I. Rodnianski, "A fully anisotropic mechanism for formation of trapped surfaces in vacuum", *Inventiones* **198** (2014), pp. 1–26.

[102] S. Chandrasekhar, *The Mathematical Theory of Black Holes*, International Series of Monographs on Physics, Oxford University Press: Oxford, 1983.

[103] E. Giorgi, "Electromagnetic-gravitational perturbations of Kerr–Newman spacetime: the Teukolsky and Regge–Wheeler equations", *J. Hyperbolic Differ. Eqts.* **19** (2022), pp. 1–139.

[104] E. Giorgi, “The Carter tensor and the physical-space analysis in perturbations of Kerr–Newman spacetime”, *J. Differ. Geom.* **127** (2024), pp. 277–371.

[105] E. Giorgi, “Boundedness and Decay for the Teukolsky System in Kerr–Newman Spacetime I: The Case $|a|,|Q| \ll M$”, preprint, 2023, 2311.07408.

[106] E. Giorgi and J. Wan, “Boundedness and Decay for the Teukolsky System in Kerr–Newman Spacetime II: The Case $|a| \ll M, |Q| < M$ in Axial Symmetry”, preprint, 2407.10750.

[107] L. He, “The linear stability of weakly charged and slowly rotating Kerr–Newman family of charged black holes”, preprint, 2023, 2301.08557. 1–180.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 151-175
https://doi.org/10.5802/crmeca.278

Review article / *Article de synthèse*

On the regularity problems of Einstein equations

Sur les problèmes de régularité pour les équations d'Einstein

Qian Wang[a]

[a] Mathematical Institute, University of Oxford, UK
E-mail: qian.wang@maths.ox.ac.uk

Abstract. In this survey, we provide a review of recent progresses in the local well-posedness problem of Einstein equations in (3+1)-D with low regularity and its applications.

Résumé. Dans cette enquête, nous rendons compte des progrès récents dans le problème de la bonne pose locale des équations d'Einstein en (3+1)-D avec une faible régularité et ses applications.

Keywords. Cauchy problem, Einstein equations, Local well-posedness.

Mots-clés. Problème de Cauchy, Équations d'Einstein, Bien-posé local.

Manuscript received 4 November 2024, revised and accepted 5 December 2024.

1. Introduction

Many fundamental theories of physics are modelled by nonlinear hyperbolic equations, such as general relativity, gas and fluid mechanics, elastodynamics etc. The main theories of these PDEs are on the well-posedness of the solution of the equation with certain initial or boundary conditions, which are usually classified into local-in-time or global-in-time well-posedness. Pioneered by Schauder [1], Choquet-Bruhat [2] and Leray [3], the local-in-time well-posedness of classical solutions for nonlinear hyperbolic equations of multiple spatial dimension were studied by physical methods such as energy methods, pointwise estimates based on Kirchhoff formulas.

In [2], without assuming analyticity for the Cauchy data, Choquet-Bruhat proved local existence result for the following second order hyperbolic system[1]

$$\begin{cases} A^{\mu\nu}(\Phi,\boldsymbol{\partial}\Phi)\partial_\mu\partial_\nu\Phi = F(\Phi,\boldsymbol{\partial}\Phi), \\ \Phi|_{t=0} = \Phi_0, \quad \partial_t\Phi|_{t=0} = \Phi_1 \end{cases} \tag{1.1}$$

where Φ is a vector-valued function, the principal coefficients $A^{\mu\nu}$ and F are smooth functions of their variables. In particular she provided an constructive proof for local existence and uniqueness of solutions for Einstein vacuum equations under wave coordinates. Together with

[1] By default $\partial_0 = \partial_t$ unless specified otherwise. Under the Einstein summation convention, the lowercase Latin indices such as i, j, k start from 1, while the Greek indices such as μ, ν begin from 0. Here, $\boldsymbol{\partial}$ represents both the spatial and time derivatives, while ∂ without a subscript is reserved solely for spatial derivatives.

Geroch, in [4], she also proved that for each initial data set, there is a unique maximal future development.

Consider Einstein vacuum equations

$$\mathbf{R}_{\alpha\beta}(\mathbf{g}) = 0 \tag{1.2}$$

where $\mathbf{R}_{\alpha\beta}$ denotes Ricci curvature tensor of the Lorentzian spacetime $(\mathscr{M}, \mathbf{g})$. An initial data set for (1.2) consists of a three dimensional surface Σ_0 together with a Riemmannian metric g and a symmetric 2-tensor k satisfying the constraint equations

$$\begin{cases} \nabla^j k_{ij} - \nabla_j \mathrm{tr} k = 0 \\ R_s - |k|^2 + (\mathrm{Tr} k)^2 = 0 \end{cases} \tag{1.3}$$

where ∇ is the Levi-Civita connection of the metric g on Σ_0, and R_s is the scalar curvature of (Σ, g). For a given initial data set (g, k, Σ_0), solving the Cauchy problem is to find a metric $\mathbf{g}$ satisfying (1.2) and an embedding of Σ_0 in $\mathscr{M}$ such that the metric induced by $\mathbf{g}$ on Σ_0 is g and the 2-tensor k is the second fundamental form of the hypersurface $\Sigma_0 \subset \mathscr{M}$.

With the wave coordinates x^α, the metric takes

$$\mathbf{g} = -n^2 \mathrm{d}t^2 + g_{ij}(\mathrm{d}x^i + v^i \mathrm{d}t)(\mathrm{d}x^j + v^j \mathrm{d}t),$$

where n is the lapse function of $x^0 = t$ and the vector-valued function v^i is the shift of the metric, g_{ij} is the induced Riemannian metric on Σ_t, the level set of t. Since $\mathbf{g}$ is Lorentzian, there is a constant $c > 0$ such that

$$c^{-1}|\xi|^2 < g_{ij}\xi^i\xi^j \le c|\xi|^2, \quad n^2 - |v|_g^2 \ge c.$$

Schematically, with $\Phi = (\mathbf{g}_{\mu\nu})$, under the wave coordinates, the reduced Einstein equation system takes the form

$$\mathbf{g}^{\mu\nu}(\Phi)\partial_\mu\partial_\nu\Phi = \mathscr{N}(\Phi, \boldsymbol{\partial}\Phi). \tag{1.4}$$

$\mathbf{g}^{\mu\nu}$ is the inverse metric of the Lorentz metric $\mathbf{g}$; the function $\mathscr{N} = (\mathscr{N}_{\mu\nu})$ on the right-hand side is smooth on its variables, and $\mathscr{N}(\Phi, \boldsymbol{\partial}\Phi)$ is quadratic in $\boldsymbol{\partial}\Phi$. The method Choquet-Bruhat adopted is based on pointwise estimates and a generalized Kirchhoff formula, which later inspired important works such as the breakdown criterion for solutions of Einstein equations given by Klainerman–Rodnianski in [5–7]. We will go back to this point in Section 5. Meanwhile, the energy method adopted in [1,3] has become the classical method in studying hyperbolic PDEs.

A key motivation of the work [2] was to extend the Cauchy–Kowalevski theorem to non-analytic Cauchy data, as the assumption of analyticity is meaningless in a physical theory where coordinate changes are only restricted to be sufficiently differentiable. Furthermore, in [2], Choquet-Bruhat noted: "It seemed to me that, for the problems considered by the theory of relativity, it would be interesting to obtain, under the minimal possible amount of assumptions, an existence theorem easy to use, enabling [one] to find properties of the solutions that can be compared with the classical properties of light waves and gravitational potentials, and to have formulas which can be an efficient method of calculating gravitational fields, at least approximately, that correspond to given initial conditions."

Since the seminal work [2] of Choquet-Bruhat, extensive work has been done on the well-posedness of the quasi-linear wave equation (1.1) in $\mathbb{R}^{n+1}$ for $n \ge 2$, including applications to Einstein-vacuum equations. The assumptions regarding the regularity of the Cauchy data have been significantly relaxed. The classical approach for proving local well-posedness relies on energy methods, Sobolev embeddings, and the classical iteration argument. Constructing solutions with lower regularity data may involve using Strichartz estimates or bilinear estimates (when the equations satisfy the null form condition). These approaches may go beyond physical methods, requiring constructing and controlling a parametrix and employing Fourier analysis, which are challenging to extend to quasilinear equations.

Next, we provide a detailed review of the local well-posedness results for Cauchy problems concerning Einstein equations and similar types of quasilinear wave equations.

The classical local well-posedness result of Hughes–Kato–Marsden [8] in the Sobolev space H^s follows from the energy estimate

$$\|\boldsymbol{\partial}\Phi(t)\|_{H^{s-1}} \le c\|\boldsymbol{\partial}\Phi(0)\|_{H^{s-1}} \cdot \exp\left(\int_0^t \|\boldsymbol{\partial}\Phi(\tau)\|_{L^\infty_x} \mathrm{d}\tau\right), \tag{1.5}$$

as well as the Sobolev embedding and a standard iteration argument. This result holds for any $s > n/2+1$, since $H^s \subset L^\infty$ gives the control of $\|\boldsymbol{\partial}\Phi\|_{L^\infty_t L^\infty_x}$.

Theorem 1 ([8, 9]). *Let (Σ_0, g, k) be an initial data set for the Einstein vacuum equation (1.2) in $(3+1)$ spacetime. Assume Σ_0 can be covered by a locally finite system of coordinate charts, related to each other by C^1 diffeomorphisms, such that $(g,k) \in H^s_{\mathrm{loc}}(\Sigma_0) \times H^{s-1}_{\mathrm{loc}}(\Sigma_0)$ with $s > 5/2$. Then there exists a unique global hyperbolic development $(\mathcal{M}, \mathbf{g})$ verifying (1.2), for which Σ_0 is a Cauchy hypersurface.*

One can refer to [10, pages 304–310] for the classical local existence result for Einstein vacuum equations with maximal foliation, which means the mean curvature of the level set of t is trivial. For Einstein vacuum equations with constant mean curvature foliation, which means the mean curvature of the level set of t equals t, with $t < 0$, by setting on the level set of t the spatial harmonic gauge, Andersson–Moncrief proved the local well-posedness of the solution in [11], for data $(g,k) \in H^s \times H^{s-1}$ with $s > 5/2$.

In pursuit of an existence theorem for the Einstein vacuum equations with minimal regularity assumptions, Klainerman proposed the bounded L^2 curvature conjecture in [12], which, in brief, asserts that the local existence and uniqueness result for the Einstein vacuum equations can be extended to Cauchy data with locally finite L^2 curvature and locally finite L^2 norm of first order covariant derivatives of k (see Theorem 3 for a more precise statement). Additionally, in connection with the global stability of Minkowski spacetime established by Christodoulou and Klainerman in their seminal work [10], Klainerman posed the following problem in [13]:

Problem 1.1 (Strong stability of Minkowski space). *Does there exists a scale invariant smallness condition such that all developments, whose initial data sets (Σ, g, k) verify it, have complete maximal future developments?*

The bounded L^2 curvature conjecture was regarded as a highly challenging problem, a key step toward resolving the above question. Significant progress was made in the early 2000s to lower the regularity assumptions for local well-posedness results of quasilinear wave equations of type (1.4).

To improve the classical results, one key objective is obtaining a better estimate on $\|\boldsymbol{\partial}\Phi\|_{L^1_t L^\infty_x}$. This boils down to deriving the Strichartz estimates for the wave operator $\mathbf{g}^{\alpha\beta}(\Phi)\partial_\alpha\partial_\beta$ which has rough coefficients, only as smooth as Φ. In this context, for $s < n/2+1$, Strichartz estimates provide a pathway to improving regularity, which is crucial for enhancing classical results. One can refer to [14–28] the works of Smith, Buhari, Chemin, Tataru, Klainerman, Rodnianki, Wang for results of $s < 5/2$ for the quasilinear wave equation of the type of (1.4). In particular Klainerman and Rodnianski gave the $s > 2$ result for Einstein vacuum equation under the wave coordinate gauge in [22–24], which is sharp due to the counter example given by Lindblad–Ettinger [29]. And the result of Smith–Tataru [28][2] and Wang [27] achieve the sharp $s > 2$ result for the general equation, (which takes the form as (1.4), but the metric does not have trivial Ricci curvature, see (2.2)). The counter example for the case of $s = 2$ is given by Lindblad [30].

[2] The result is for the dimension $2 \le n \le 5$, at least sharp in $n = 2, 3$.

Nevertheless, Strichartz estimates are established for the wave operator. When the nonlinearity $\mathcal{N}(\Phi, \boldsymbol{\partial}\Phi)$ exhibits a more delicate structure, better results are expected, which cannot be achieved solely through Strichartz estimates. For instance, for semilinear wave equations in $\mathbb{R}^{3+1}$

$$\Box\phi = \mathcal{N}(\phi, \boldsymbol{\partial}\phi), \quad \phi(0) = \phi_0, \partial_t\phi(0) = \phi_1,$$

using Strichartz estimate, the local-wellposedness was proved for $(\phi_0, \phi_1) \in H^s \times H^{s-1}$ with $s > 2$ by Ponce–Sideris [31] and the result is sharp due to the counter example by Lindblad [32]. However if the quadratic forms are null forms (see (3.23)), the result is improved by Klainerman and Machedon in [33] to $s = 2$ by establishing bilinear estimates.

For geometric wave equations, those are semilinear and exhibiting null conditions, including wave map, Maxwell–Klein–Gordon and Yang–Mills equations in Minkowski space, the study of low-regularity solutions is driven by investigating global-in-time behavior. An important example is the global finite energy solution for Maxwell–Klein–Gordon equations in Minkowski space, established in [34], as well as for Yang–Mills equations in [35]. For both cases, the energy of the curvature (together with the scalar field for Maxwell–Klein–Gordon equations) is conserved over time. In $\mathbb{R}^{3+1}$, Klainerman and Machedon [34, 35] proved in the 1990s that local solutions can be constructed for initial data with finite energy. The conservation of energy allows these local solutions to be extended globally in time. Their results were based on innovative bilinear estimates for null forms, improving classical results by more than half a derivative.

Further progress on well-posedness for lower regularity solutions was made by employing bilinear estimates in the wave-Sobolev space, developed independently by Klainerman–Machedon and Bourgain (see [33, 36, 37]). For instance, the scale-invariant, optimal regularity result on Maxwell–Klein–Gordon equations in $\mathbb{R}^{4+1}$ was achieved by Lührmann–Krieger [38] and Oh–Tataru [39], where global well-posedness for finite energy data was proven; one can refer to Krieger–Tataru [40] and Oh–Tataru [41] for the optimal regularity result in $\mathbb{R}^{4+1}$ for Yang–Mills equations.

While bilinear estimates rely heavily on Fourier analysis and the null form condition, extending these estimates to quasilinear wave equations poses a fundamental challenge. In the case of Einstein equations with maximal foliation, by using the Yang–Mills formalism and adopting the spatial Coulomb gauge, the equations exhibit the structure of null forms. The $s = 2$ result for Einstein vacuum equations is established by Klainerman, Rodnianski and Szeftel in [42] and [43–46] by taking advantage of this structure, which is so far the best regularity result for Einstein vacuum equations. The counter example in [29] implies that, under wave coordinate gauge, the so-called weak null form observed by Lindblad–Rodnianski [47] is insufficient to yield the result of the bounded L^2 curvature conjecture. In [42], it is suggested that, although the result is not optimal with respect to the standard scaling of the Einstein equation, this result may be sharp due to "null scaling", which is related to establishing a lower bound on the null radius of injectivity. This bound is crucial for their parametrix construction and establishing bilinear and trilinear estimates. To control the lower bound, it relies on a series of sharp trace estimates along null hypersurfaces. We will review this set of estimates in Sections 4 and 5.

In Section 2, we review the $s > 2$ result for the Einstein vacuum equations and similar types of equations. Section 3 provides a brief overview of work on resolving the bounded L^2 curvature conjecture. In Section 4, we discuss methods for controlling causal geometry in rough Einstein spacetimes. Finally, as applications of controlling null hypersurfaces with limited regularity, we review works on breakdown criteria for the Einstein equations in Section 5.

2. Rough solutions for Einstein equations

In the section, we review the result of Klainerman and Rodnianski in [22–24].

Theorem 2. *Consider the classical solution of (1.4), the reduced Einstein equation under the wave coordinates. Suppose* $\mathbf{g}(0)$ *is a continuous Lorentz metric, and* $\sup_{|x|=r}|\mathbf{g}_{\alpha\beta}(0)-\mathbf{m}_{\alpha\beta}|\to 0$ *as* $r\to\infty$. *The time* T *of the existence depends only on the size of the form* $\|\boldsymbol{\partial}\mathbf{g}_{\mu\nu}(0)\|_{H^{s-1}}$ *for any fixed* $s>2$.

One can refer to the counter example given in [29], which shows that Theorem 2 is sharp. The above result is extended to Einstein vacuum equation with CMC spatial harmonic gauge by Wang in [25, 26].

Since one can approximate a given H^s initial data set for the Einstein vacuum equations by classical initial data sets, i.e. $H^{s'}$ data sets with $s'>5/2$ for which the classical solutions exist and unique. The above theorem allows one to pass to the limit and derive existence of solutions for the given, rough, initial data set. In particular, one can refer to the result of Maxwell (which is for $s>3/2$) in [48] for justifying such approximation for the solutions of the constraint equations at the initial slice.

Recall the energy estimates for the solution of (1.4) in $n+1$ spacetime. To achieve the result for $s>2$, one need to control $\exp(\int_0^t\|\boldsymbol{\partial}\Phi(\tau)\|_{L^\infty_x})$. In terms of Sobolev embedding

$$\|\boldsymbol{\partial}\Phi\|_{L^\infty_x}\lesssim\|\boldsymbol{\partial}\Phi\|_{H^s_x},\quad s>\frac{n}{2}$$

which is consistent with the classical result. For the solution of the linear wave equation $\Box\phi=0$ in Minkowski spacetime $\mathbb{R}^{n+1}$, there holds the Strichartz estimate

$$\|\boldsymbol{\partial}\phi\|_{L^2_{[0,T]}L^\infty_x}\le cT^\sigma(\|\phi_0\|_{H^{\frac{n}{2}+\frac{1}{2}+\sigma}}+\|\phi_1\|_{H^{\frac{n}{2}-\frac{1}{2}+\sigma}}) \tag{2.1}$$

with $\sigma>0$ arbitrarily small.

Consider the following quasilinear wave equation, similar in form to (1.4),

$$\begin{cases}-\partial_t^2\phi+g^{ij}(\phi)\partial_i\partial_j\phi=\mathcal{N}(\phi,\boldsymbol{\partial}\phi),\\ \phi|_{t=0}=\phi_0,\quad \partial_t\phi|_{t=0}=\phi_1\end{cases} \tag{2.2}$$

with g^{ij} bounded, uniformly elliptic and smooth about its variable. To achieve a better result than the classical one, it is important to recover the Strichartz estimate in (2.1) in the rough, curved spacetime with the metric $\mathbf{g}=-\mathrm{d}t^2+g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ depending on the solution. Strichartz estimate for wave equations with rough coefficients was first studied by Smith [14]. The breakthrough was then made by Bahouri–Chemin [18, 19] and by Tataru [15] using parametrix constructions. By establishing a Strichartz estimate for solutions to the linearized equation $-\partial_t^2\phi+g^{ij}(\phi)\partial_i\partial_j\phi=0$ of the form

$$\|\boldsymbol{\partial}\phi\|_{L^2_tL^\infty_x}\le c\left(\|\phi_0\|_{H^{2+\alpha}}+\|\phi_1\|_{H^{1+\alpha}}\right)$$

with a loss of $\alpha>1/4$, they obtained the well-posedness of (2.2) in H^s with $s>2+1/4$. This result was later improved to $s>2+1/6$ in [17].

An important next step was taken in [20] and [21], where in [20] Klainerman introduced the commuting vectorfields approach for Strichartz estimates, with which he reproduced the $s>2+1/6$ result. In particular he pointed out the regularity of null hypersurface can be improved by adopting the following decomposition for $\mathbf{R}_{LL}$, the component of Ricci curvature of the metric $\mathbf{g}$ contracted by the null vector-field L,

$$\mathbf{R}_{LL}=LP-\tfrac{1}{2}L^\mu L^\nu\mathbf{g}^{\alpha\beta}\partial_\alpha\partial_\beta\mathbf{g}_{\mu\nu}+E \tag{2.3}$$

with P roughly $\mathbf{g}\cdot\boldsymbol{\partial}\mathbf{g}$ and $E=\mathbf{g}(\boldsymbol{\partial}\mathbf{g})^2$. This important geometric treatment introduced in [20] is crucially used in [21]. In [21], blended with the paradifferential calculus ideas initiated in [19], [16] and [17], Klainerman and Rodnianski improved the local well-posedness of (2.2) in $\mathbb{R}^{1+3}$ to data in the Sobolev space H^s with $s>2+(2-\sqrt{3})/2$. According to the counter-examples in [30], one can only expect to obtain the local well-posedness in H^s with $s>2$. Note that the metric of (2.2) is more general in the sense that it does not verify Einstein vacuum equation (1.2).

For (2.2), the optimal $s > 2$ result is achieved in [28] and [27], both of which take advantage of the decomposition (2.3) to control the causal geometry.

Next we briefly summarize the main reduction steps for the proof of Theorem 2.

To prove Theorem 2, under the bootstrap assumption

$$\|\boldsymbol{\partial}\Phi\|_{L^2_{[0,T]}L^\infty_x} \le 1, \tag{2.4}$$

one needs to prove the energy estimate and Strichartz estimate below

$$\|\boldsymbol{\partial}\Phi\|_{L^\infty_{[0,T]}H^{1+\gamma}} \lesssim \|\boldsymbol{\partial}\Phi(0)\|_{H^{1+\gamma}}, \quad \|\boldsymbol{\partial}\Phi\|_{L^2_{[0,T]}L^\infty_x} \le CT^\delta, \tag{2.5}$$

with some $\delta > 0$, $C > 0$ a universal constant depending on the bound of $\|\boldsymbol{\partial}\Phi(0)\|_{H^{1+\gamma}}$. The energy estimate follows from (1.5) by using (2.4). Assuming the Strichartz estimate in (2.5), with T sufficiently small, (2.4) is improved.

Step 1. Reduction to dyadic Strichartz estimate. To prove the Strichartz estimate in (2.5), it suffices to prove

$$\|P_\lambda \boldsymbol{\partial}\Phi\|_{L^2_{[0,T]}L^\infty_x} \lesssim c_\lambda T^\delta \|\boldsymbol{\partial}\Phi(0)\|_{H^{1+\gamma}} \tag{2.6}$$

where P_λ is the Littlewood–Paley projection with the frequency $\lambda \in 2^{\mathbb{Z}}$, and $\sum_\lambda c_\lambda \le 1$. In fact, it suffices to focus on a fixed dyadic frequency with $\lambda > \Lambda_0$ with Λ_0 sufficiently large.

With ϵ_0 fixed and $0 < \epsilon_0 < \gamma/5$, divide $[0,T]$ into subintervals I with size $\approx T\lambda^{-8\epsilon_0}$. On an interval I consider the linearized equation

$$\mathbf{g}^{\alpha\beta}_{<\lambda}\partial_\alpha\partial_\beta\psi = 0 \tag{2.7}$$

where $\mathbf{g}^{\alpha\beta}_{<\lambda} = \sum_{0<\mu\le 2^{-M_0}\lambda} P_\mu \mathbf{g}^{\alpha\beta}$, with $M_0 > 0$ a fixed large constant.

For the frequency localized data, satisfying

$$(2^{-10}\lambda)^m \|\boldsymbol{\partial}\psi(0)\|_{L^2_x} \le \|\nabla^m \boldsymbol{\partial}\psi(0)\|_{L^2_x} \le (2^{10}\lambda)^m \|\boldsymbol{\partial}\psi(0)\|_{L^2_x}$$

prove that with $\delta > 0$ such that $5\epsilon_0 + \delta < \gamma$ there holds

$$\|P_\lambda \boldsymbol{\partial}\psi\|_{L^2_I L^\infty_x} \lesssim |I|^\delta \|\boldsymbol{\partial}\psi(0)\|_{\dot{H}^{1+\delta}}. \tag{2.8}$$

Step 2. Rescaling. With the metric

$$H_{(\lambda)}(t,x) = \mathbf{g}_{<\lambda}(\lambda^{-1}t, \lambda^{-1}x)$$

consider the rescaled equation of (2.7)

$$H^{\alpha\beta}_{(\lambda)}(t,x)\partial_\alpha\partial_\beta\psi = 0$$

in the region $[0,t_*] \times \mathbb{R}^3$, where $t_* \le \lambda^{1-8\epsilon_0}$. Then (2.8) can be obtained by showing

$$\|P\boldsymbol{\partial}\psi\|_{L^2_{[0,t_*]}L^\infty_x} \lesssim t_*^\delta \|\boldsymbol{\partial}\psi(0)\|_{L^2_x},$$

where $P = P_1$.

Step 3. Reduction to an $L^1 - L^\infty$ decay estimate. By running a TT^* type argument, the proof of the dyadic strichartz estimate in Step 2 can be further reduced to an $L^1 - L^\infty$ decay estimate. Let ψ be the solution of

$$\Box_{H_{(\lambda)}}\psi = 0 \tag{2.9}$$

with data at $t = t_0 \in [0,t_*]$. Prove

$$\|P\boldsymbol{\partial}\psi\|_{L^\infty_x} \lesssim \left(\frac{1}{(1+|t-t_0|)^{1-\delta}} + d(t)\right)\sum_{k=0}^m \|\nabla^k \boldsymbol{\partial}\psi(t_0)\|_{L^1_x} \tag{2.10}$$

for some integer $m > 0$, where $\delta > 0$ is sufficiently small, and $t_*^{1/q}\|d\|_{L^q([0,t_*])} \lesssim 1$ for some $q > 2$ sufficiently close to 2.

Step 4. Reduction to a localized $L^2 - L^\infty$ estimate. The next step is to further localize in space and reduce (2.10) to the following result

$$\|P\boldsymbol{\partial}\psi(t)\|_{L^\infty_x} \lesssim \left(\frac{1}{(1+|t-t_0|)^{2/q}} + d(t)\right)\sum_{k=0}^{m-2}\|\nabla^k\boldsymbol{\partial}\psi(t_0)\|_{L^2_x}, \tag{2.11}$$

where ψ is the solution of (2.9) and with data supported within a geodesic ball of radius 1/2, centered at the origin at Σ_{t_0}.

In the spacetime slab $[0,t_*]\times\mathbb{R}^3$, let u be the optical function satisfying $H^{\alpha\beta}_{(\lambda)}\partial_\alpha u\partial_b u = 0$, with initial condition properly chosen, and $\underline{u} = 2t - u$. Let

$$L' = -H^{\alpha\beta}_{(\lambda)}\partial_\beta u\partial_\alpha, \quad \mathbf{b}^{-1} = -\langle L', \mathbf{T}\rangle = \mathbf{T}(u).$$

Define the canonical null pair L and $\underline{L}$ by

$$L = \mathbf{b}L', \quad \underline{L} = 2\mathbf{T} - L,$$

where $\mathbf{T}$ is the future-directed time-like unit normal of Σ_t in the spacetime. Let

$$Q_{\alpha\beta}[f] = \partial_\alpha f\partial_\beta f - \tfrac{1}{2}H_{(\lambda)\alpha\beta}H^{\mu\nu}_{(\lambda)}\partial_\mu f\partial_\nu f.$$

It is then reduced to bound the conformal energy, defined with the energy density constructed by contracting the standard current $\mathscr{P}^\alpha[\psi] = Q_{\alpha\beta}[\psi]\mathbf{T}^\beta$ with the Morawetz vector field $K = (1/2)n(u^2\underline{L} + \underline{u}^2 L)$, followed with a proper normalization.

To control the propagation of such energy, it is important to control the deformation tensor of K, defined by ${}^{(K)}\pi_{\alpha\beta} = (\mathbf{D}_\alpha K)_\beta + (\mathbf{D}_\beta K)_\alpha$. As a result of rescaling, the original spacetime slab stretches by the high frequency $\lambda > \Lambda_0$, so the analysis is carried out in a spacetime with a metric close to Minkowski. The derivatives of the mollified metric $H_{(\lambda)}$ are bounded by powers of the frequency λ. Since K is conformal killing in Minkowskian space, it is expected to be nearly conformal killing in the spacetime slab with the metric $H_{(\lambda)}$, with the error displaying smallness in the form of negative powers of λ.

The level surfaces of u, denoted by $\mathscr{H}_u$, are the outgoing lightcones. Using an arbitrary orthonormal frame $(e_A)^2_{A=1}$ on $S_{t,u} = \Sigma_t \cap \mathscr{H}_u$, define the connection coefficients

$$\begin{aligned} \chi_{AB} &= \langle \mathbf{D}_A L, e_B\rangle, \quad \underline{\chi}_{AB} = \langle \mathbf{D}_A\underline{L}, e_B\rangle \\ \zeta_A &= \tfrac{1}{2}\langle \mathbf{D}_{\underline{L}}L, e_A\rangle, \quad \underline{\eta}_A = \tfrac{1}{2}\langle \mathbf{D}_L\underline{L}, e_A\rangle. \end{aligned} \tag{2.12}$$

It is necessary to obtain sufficient decay in terms of λ for $\mathrm{tr}\chi - 2/n(t-u), \hat{\chi}, \zeta, \underline{\eta}$ and for the first order derivatives of $\mathrm{tr}\chi$. This can be accomplished by using a set of null structure equations, such as the Raychaudhuri equation:

$$L\mathrm{tr}\chi + \tfrac{1}{2}(\mathrm{tr}\chi)^2 = -|\hat{\chi}|^2 - \mathbf{R}_{LL}(H_{(\lambda)}) + \cdots$$

However, applying (2.3) with metric $H_{(\lambda)}$ is insufficient to give the desired bound on the first order derivatives of $\mathrm{tr}\chi$. To improve the result of Klainerman–Rodnianski in [21] to the $s > 2$ result in Theorem 2, it is crucial to use the Einstein vacuum equation (1.2) to achieve sufficient decay for the connection coefficients. To this end, it is shown in [24] that the Ricci tensor of $H_{(\lambda)}$ and its derivatives are close to zero in a certain sense, given that the original metric is Einstein vacuum.

To avoid establishing the delicate comparison estimates in [24], [25] adapts the reduction procedure to bound the conformal energy of the lowest order within the domain of influence of a unit ball in the original metric. The difficulty is reduced to obtain the necessary control of the connection coefficients in the rough Einstein spacetime. This set of estimates is derived by directly taking advantage of the geometric structure of Einstein vacuum spacetime in [26]. Additional challenges arise in the CMCSH gauge, where certain estimates, such as the time derivative of the shift of the metric cannot be obtained. It is crucial in particular for providing sufficient control of the null cones in the rough spacetime. This issue is circumvented by normalizing the basic energy current.

The adapted reduction procedure is used in [27] to prove the sharp $s > 2$ result for the general equation (2.2). In [27] the construction of conformal energy is further adapted by using the method of Dafermos–Rodnianski in [49], which additionally achieves the crucial conformal flux. The added control relaxes slightly the requirements on the causal geometry in the rough spacetime. By using the conformal invariance of the null hypersurface, introduce a conformal change of the spacetime metric with a well chosen conformal factor. Then with respect to the normalized metric, the required control on the causal geometry to bound the newly constructed conformal energy can be obtained in the rough spacetime. In [50], Wang further adapts the methods in [25,27] to give the low-regularity well-posedness result for compressible Euler equation in 3-D.

3. The bounded L^2 curvature conjecture for Einstein vacuum equations

The bounded L^2 curvature conjecture for Einstein vacuum equations in $(3+1)$ spacetime is resolved in [42] and [43–46]. The conjecture states that

Theorem 3 ([42]). *Let $(\mathscr{M}, \mathbf{g})$ be an asymptotically flat solution to the Einstein vacuum equations (1.2) together with a maximal foliation by space-like hypersurfaces Σ_t defined as level hypersurfaces of a time function t. Assume that the initial slice $(\Sigma_0, \mathbf{g}, k)$ is such that the Ricci curvature $Ric, \nabla k \in L^2(\Sigma_0)$ and Σ_0 has a strictly positive volume radius on scales ≤ 1, i.e. $r_{\mathrm{vol}}(\Sigma_0, 1) > 0$. Then, there exists a time $T = T(\|Ric\|_{L^2(\Sigma_0)}, \|\nabla k\|_{L^2(\Sigma_0)}, r_{\mathrm{vol}}(\Sigma_0, 1)) > 0$ and a constant $C = C(\|Ric\|_{L^2(\Sigma_0)}, \|\nabla k\|_{L^2(\Sigma_0)}, r_{\mathrm{vol}}(\Sigma_0, 1)) > 0$ such that the following control holds on $0 \leq t \leq T$:*

$$\|\mathbf{R}, \nabla k\|_{L^\infty_{[0,T]} L^2_{\Sigma_t}} \leq C, \quad \textit{and} \quad \inf_{0 \leq t \leq T} r_{\mathrm{vol}}(\Sigma_t, 1) \geq 1/C.$$

The above result can be regarded as a continuation principle for Einstein vacuum equation under the maximal foliation gauge. Moreover, the above result is used by Czimek and Graf in [51] to give the low regularity existence result for the spacelike-characteristic initial data problem in Einstein spacetime.

Next we summarize the main idea and the main steps of the proof.

3.1. *The Yang–Mills formalism*

Let $\{e_\alpha\}_{\alpha=0}^3$ be an orthonormal frame on $\mathscr{M}$, i.e.

$$\mathbf{g}(e_\alpha, e_\beta) = \mathbf{m}_{\alpha\beta} = \mathrm{diag}(-1, 1, \ldots, 1),$$

with $e_0 = \mathbf{T}$. Consistent with Cartan formalism, define the connection 1-form

$$\mathbf{A}_{\alpha\beta}(X) = \mathbf{g}(\mathbf{D}_X e_\beta, e_\alpha)$$

where X is an arbitrary vectorfield in $T\mathscr{M}$. Then

$$\mathbf{R}(e_\alpha, e_\beta, \partial_\mu, \partial_\nu) = \partial_\mu(\mathbf{A}_\nu)_{\alpha\beta} - \partial_\nu(\mathbf{A}_\mu)_{\alpha\beta} + [\mathbf{A}_\mu, \mathbf{A}_\nu]_{\alpha\beta} \tag{3.1}$$

where

$$([\mathbf{A}_\mu, \mathbf{A}_\nu])_{\alpha\beta} = (\mathbf{A}_\mu)_\alpha{}^\gamma (\mathbf{A}_\nu)_{\gamma\beta} - (\mathbf{A}_\nu)_\alpha^\gamma (\mathbf{A}_\mu)_{\gamma\beta}.$$

Denote

$$(\mathbf{F}_{\mu\nu})_{\alpha\beta} = \mathbf{R}_{\alpha\beta\mu\nu}.$$

Using (1.2), it follows from the Bianchi identity that

$$\mathbf{D}^\mu \mathbf{F}_{\mu\nu} + [\mathbf{A}^\mu, \mathbf{F}_{\mu\nu}] = 0.$$

It then follows by using the above identity, (1.2), (3.1) and the fact that $\partial_\mu(\mathbf{A}_\nu) - \partial_\nu(\mathbf{A}_\mu) = \mathbf{D}_\mu \mathbf{A}_\nu - \mathbf{D}_\nu \mathbf{A}_\mu$ that

$$\Box_{\mathbf{g}} \mathbf{A}_\nu - \mathbf{D}_\nu(\mathbf{D}^\mu \mathbf{A}_\mu) = \mathbf{J}_\nu \tag{3.2}$$

where

$$\mathbf{J}_\nu = \mathbf{D}^\mu([\mathbf{A}_\mu, \mathbf{A}_\nu]) - [\mathbf{A}^\mu, \mathbf{F}_{\mu\nu}],$$

and it is direct to check that $\mathbf{D}_\mu \mathbf{J}^\mu = 0$.

Let $A_0 = \mathbf{A}_\mu e_0^\mu$ and $A_i = \mathbf{A}_\mu e_i^\mu$. It is straightforward to see that

$$\begin{aligned}(A_i)_{0j} &= (A_j)_{0i} = -k_{ij}, \quad i,j = 1,2,3\\ (A_0)_{0i} &= -n^{-1}\nabla_i n, \quad i = 1,2,3.\end{aligned}$$

Note that there is freedom to choose a frame e_1, e_2, e_3 such that the corresponding connection A satisfies the Coulomb gauge condition (see [42, Lemma 4.3])

$$\nabla^i (A_i)_{jk} = 0.$$

With $\boldsymbol{\partial} = (\partial_0, \partial)^3$ and $\mathbf{A} = (A_0, A_i)$, (3.2) is then reduced to

$$\Delta A_0 = \mathbf{A}\boldsymbol{\partial}\mathbf{A} + \mathbf{A}\partial A_0 + \mathbf{A}^3 \tag{3.3}$$

$$\Box_{\mathbf{g}} A_i + \partial_i(\partial_0 A_0) = A^j \partial_j A_i + A^j \partial_i A_j + A_0 \boldsymbol{\partial}\mathbf{A} + A\boldsymbol{\partial} A_0 + \mathbf{A}^3. \tag{3.4}$$

The spatial derivative estimates of A_0 can be obtained by using elliptic estimates and (3.3). It is more important to control $A = A_i$, for which one may need to employ the wave equation (3.4). For this purpose, one needs eliminate the term $\partial_i(\partial_0 A_0)$ in (3.4), which would be done by projecting the equation onto divergence free vector-fields with the help of a non-local operator if following the treatment in [35]. However, adapting this approach to the Einstein vacuum equations is too complicated. In [42], by introducing

$$B_i = (-\Delta)^{-1}(\mathrm{curl}\,(A)_i), \tag{3.5}$$

instead of directly using (3.4) to control A, the problem reduces to controlling ∂B, based on the comparison estimate [42, Lemma 6.5]

$$A = \mathrm{curl}\, B + E \tag{3.6}$$

where E is a controllable error. In [42], the authors rely on $\Box_{\mathbf{g}} B$ to obtain derivative estimates of B.

3.2. *Bilinear and Trilinear estimates*

The proof of Theorem 3 relies on bilinear and trilinear estimates, which will be outlined in this subsection.

The first step is to reduce the proof of Theorem 3 to a small data problem (see [42, Section 2.3]). Under the assumption in Theorem 3, one can use [52, 53] to obtain the lower bound of harmonic radius depending on $\|R\|_{L^2(\Sigma_0)}$ and the lower bound of initial volume radius. Within the geodesic ball of radius no more than the lower bound, there are coordinates with respect to which the metric is comparable to Euclidean metric and is bounded in H^2. Using this result, by localizing, rescaling and adapting the gluing process in [54, 55], Theorem 3 is reduced to

Theorem 4 ([42, Theorem 2.10]). *Let $(\mathscr{M}, \mathbf{g})$ be an asymptotically flat solution to the Einstein vacuum equations (1.2) together with a maximal foliation by space-like hypersurfaces Σ_t defined as level hypersurfaces of a time function t. Assume that the initial slice (Σ_0, g, k) is such that:*

$$\|R\|_{L^2(\Sigma_0)} \le \varepsilon, \quad \|k, \nabla k\|_{L^2(\Sigma_0)} \le \varepsilon, \quad r_{\mathrm{vol}}(\Sigma_0, 1) \ge \tfrac{1}{2},$$

then there exists a small universal constant $\varepsilon_0 > 0$ such that if $0 < \varepsilon < \varepsilon_0$, the following control holds on $0 \le t \le 1$

$$\|\mathbf{R}, k, \nabla k\|_{L^\infty_{[0,1]} L^2(\Sigma_t)} \lesssim \varepsilon, \quad r_{\mathrm{vol}}(\Sigma_t, 1) \ge \tfrac{1}{4}.$$

[3] $\partial_0 = \partial_{e_0}, \partial_i = \partial_{e_i}$, and denote for simplicity $\partial = \partial_i$.

Theorem 4 is proved by using a bootstrap argument.

Assumption 3.1. *Let $M \geq 1$ be a large enough constant, to be chosen later in terms only of universal constants. By choosing $\varepsilon > 0$ small, $M\varepsilon$ can be small enough.*

Assume in the spacetime slab $\cup_{t\in[0,1]}\Sigma_t$,

$$\|\mathbf{R}\|_{L_t^\infty L^2(\Sigma_t)} \leq M\varepsilon, \quad \|\mathbf{R}\cdot L\|_{L^2(\mathscr{H})} \leq M\varepsilon \tag{3.7}$$

$$\|A\|_{L_t^\infty L^2(\Sigma_t)} + \|\boldsymbol{\partial} A\|_{L_t^\infty L^2(\Sigma_t)} \leq M\varepsilon \tag{3.8}$$

where $\mathscr{H}$ denotes the null hypersurface in the spacetime slab, with future directed normal L, normalized by $\langle L, \mathbf{T}\rangle = -1$; assume the following bilinear and trilinear estimates,

$$\|k_{ij}\partial^i\phi\|_{L^2(\mathscr{M})} \leq M^2\varepsilon \sup_{\mathscr{H}} \|\nabla\phi\|_{L^2(\mathscr{H})} + M\varepsilon\|\partial\phi\|_{L_t^\infty L^2(\Sigma_t)} \tag{3.9}$$

$$\left|\int_{\mathscr{M}} Q_{ij\gamma\delta} k^{ij} e_0^\gamma e_0^\delta \mathrm{d}\mathscr{M}\right| \leq \mathscr{M}^4\varepsilon^3 \tag{3.10}$$

with $Q_{ij\gamma\delta}$ components of the Bel–Robinson tensor; and assume a set of estimates on A_0 and its derivatives in the spacetime slab. (See the full set of bootstrap assumptions in [42, Section 5.3].)

Next we follow the steps given in [42] to improve the bootstrap assumptions.

3.2.1. *Improvement of (3.7)*

Under the bootstrap assumptions, one can obtain by elliptic estimates that

$$\|n-1, \nabla n\|_{L^\infty(\mathscr{M})} \lesssim M\varepsilon. \tag{3.11}$$

To improve (3.7), one may use the Bel–Robinson tensor

$$Q_{\alpha\beta\gamma\delta} = \mathbf{R}_\alpha{}^\lambda{}_\gamma{}^\delta \mathbf{R}_{\beta\lambda\delta\sigma} + {}^*\mathbf{R}_\alpha{}^\lambda{}_\gamma{}^{\sigma *}\mathbf{R}_{\beta\lambda\delta\sigma}.$$

Let $P_\alpha = Q_{\alpha\beta\gamma\delta} e_0^\beta e_0^\gamma e_0^\delta$. Then

$$\mathbf{D}^\alpha P_\alpha = 3Q_{\alpha\beta\gamma\delta}\pi^{\alpha\beta} e_0^\gamma e_0^\delta, \tag{3.12}$$

where $\pi_{\alpha\beta} = \mathbf{D}_\alpha \mathbf{T}_\beta + \mathbf{D}_\beta \mathbf{T}_\alpha$. With $h_{\alpha\beta} = \mathbf{g}_{\alpha\beta} + 2(e_0)_\alpha (e_0)_\beta$, define the norm $|\cdot|$ for a spacetime tensor U by

$$|U|^2 = U_{\alpha_1\cdots\alpha_k} U_{\alpha_1'\cdots\alpha_k'} h^{\alpha_1\alpha_1'} \cdots h^{\alpha_k\alpha_k'}.$$

Note the following standard property of Bel–Robinson tensor

$$P_\alpha e_0^\alpha = |\mathbf{R}|^2, \quad P_\alpha L^\alpha = |\mathbf{R}\cdot L|^2.$$

Integrating (3.12) over a spacetime region, bounded by Σ_0, Σ_t and $\mathscr{H}$, it is direct to obtain

$$\int_{\Sigma_t} |\mathbf{R}|^2 + \int_{\mathscr{H}} |\mathbf{R}\cdot L|^2 \lesssim \|\mathbf{R}\|^2_{L^2(\Sigma_0)} + \left|\int_{\mathscr{M}} Q_{\alpha\beta\gamma\delta}\pi^{\alpha\beta} e_0^\gamma e_0^\delta\right| \lesssim \varepsilon^2 + \left|\int_{\mathscr{M}} Q_{\alpha\beta\gamma\delta}\pi^{\alpha\beta} e_0^\gamma e_0^\delta\right|. \tag{3.13}$$

Noting that the nontrivial components of $\pi_{\alpha\beta}$ are

$$\pi_{ij} = -2k_{ij}, \quad \pi_{0i} = n^{-1}\nabla_i n, \mathrm{tr}k = 0, \tag{3.14}$$

the last term on the right-hand side of (3.13) can be treated by using (3.10) and (3.11)

$$\begin{aligned}\left|\int_{\mathscr{M}} Q_{\alpha\beta\gamma\delta}\pi^{\alpha\beta} e_0^\gamma e_0^\delta\right| &\lesssim \left|\int_{\mathscr{M}} Q_{ij\gamma\delta} k^{ij} e_0^\gamma e_0^\delta\right| + \left|\int_{\mathscr{M}} Q_{0i\gamma\delta} n^{-1}\nabla_i n e_0^\gamma e_0^\delta\right| \\ &\lesssim M^4\varepsilon^3 + \|\nabla n\|_{L^\infty}\|\mathbf{R}\|^2_{L_t^\infty L^2(\Sigma_t)} \\ &\lesssim M^4\varepsilon^3 + (M\varepsilon)^3.\end{aligned} \tag{3.15}$$

Substituting the above estimate into (3.13) yields

$$\int_{\Sigma_t} |\mathbf{R}|^2 + \int_{\mathscr{H}} |\mathbf{R}\cdot L|^2 \lesssim \varepsilon^2 + M^4\varepsilon^3 + (M\varepsilon)^3. \tag{3.16}$$

Here we remark that schematically

$$\left|\int_{\mathcal{M}} Q_{ij\gamma\delta} k^{ij} e_0^\gamma e_0^\delta\right| \approx \left|\int_{\mathcal{M}} k\mathbf{R}^2\right|.$$

With $\|k\|_{L_t^1 L_x^\infty}$ bounded, we can control the right-hand side by

$$\left|\int_{\mathcal{M}} Q_{ij\gamma\delta} k^{ij} e_0^\gamma e_0^\delta\right| \lesssim \|k\|_{L_t^1 L_x^\infty} \|\mathbf{R}\|^2_{L_t^\infty L^2(\Sigma_t)}.$$

Nevertheless one would not expect to control $\|k\|_{L_t^1 L_x^\infty}$, since this requires the initial data to satisfy $(g,k) \in H^{2+\epsilon}(\Sigma_0) \times H^{1+\epsilon}(\Sigma_0)$. Therefore it is crucial to rely on the assumption (3.10) instead.

3.2.2. *Improvement of (3.8)*

Next we discuss the improvement over (3.8), which is reduced to control derivatives of B, based on the following result

Proposition 5 ([42, Proposition 4.4, Lemma 6.5 and Section 11.1.1]). *The error E in (3.6) satisfies*

$$\|\partial E\|_{L_t^\infty L^2(\Sigma_t)} + \|E\|_{L_t^2 L^\infty(\Sigma_t)} \lesssim M^2\varepsilon^2. \tag{3.17}$$

There hold

$$\Box_{\mathbf{g}} B = F \tag{3.18}$$

with

$$F = (-\Delta)^{-1}[\Box_{\mathbf{g}}, \Delta]B + (-\Delta)^{-1}\Box_{\mathbf{g}}(\operatorname{curl} A),$$

and

$$\|A, \partial B\|_{L^2(\Sigma_0)} + \|\partial(\boldsymbol{\partial} B)\|_{L^2(\Sigma_0)} \lesssim \varepsilon, \quad \|\partial F\|_{L^2(\mathcal{M})} \lesssim M^2\varepsilon^2. \tag{3.19}$$

We will treat (3.18) by using the following set of estimates

Lemma 6. *Let F be a scalar function on $\mathcal{M}$, and let ϕ_0 and ϕ_1 be two scalar functions on Σ_0. Let ϕ be the solution of the following wave equation on $\mathcal{M}$:*

$$\begin{cases} \Box_{\mathbf{g}}\phi = F, \\ \phi|_{\Sigma_0} = \phi_0, \quad \partial_t\phi|_{\Sigma_0} = \phi_1. \end{cases} \tag{3.20}$$

Then ϕ satisfies the following energy estimate

$$\begin{aligned} \|\boldsymbol{\partial}\phi\|_{L_t^\infty L^2(\Sigma_t)} &+ \sup_{\mathcal{H}}(\|\nabla\!\!\!/\phi\|_{L^2(\mathcal{H})} + \|L\phi\|_{L^2(\mathcal{H})}) \\ &\lesssim \|\nabla\phi_0\|_{L^2(\Sigma_0)} + \|\phi_1\|_{L^2(\Sigma_0)} + \|F\|_{L^2(\mathcal{M})}, \end{aligned} \tag{3.21}$$

where $\mathcal{H}$ denotes null hypersurface with future directed normal L such that $\langle L, \mathbf{T}\rangle = -1$, $\nabla\!\!\!/$ denotes the Levi-Civita connection on $S_t = \mathcal{H} \cap \Sigma_t$ with respect to the induced metric γ. There also hold the higher order estimates,

$$\begin{aligned} &\|\partial\boldsymbol{\partial}\phi\|_{L_t^\infty L^2(\Sigma_t)} + \|\partial_0^2\phi\|_{L^2(\mathcal{M})} + \sup_{\mathcal{H}}\left(\|\nabla\!\!\!/(\partial\phi)\|_{L^2(\mathcal{H})} + \|L(\partial\phi)\|_{L^2(\mathcal{H})}\right) \\ &\lesssim \|\nabla^2\phi_0\|_{L^2(\Sigma_0)} + \|\nabla\phi_1\|_{L^2(\Sigma_0)} + \|\nabla F\|_{L^2(\mathcal{M})}; \end{aligned}$$

and

$$\|\Box_{\mathbf{g}}\partial\phi\|_{L^2(\mathcal{M})} \lesssim M\varepsilon(\|\nabla^2\phi_0\|_{L^2(\Sigma_0)} + \|\nabla\phi_1\|_{L^2(\Sigma_0)} + \|\nabla F\|_{L^2(\mathcal{M})}).$$

Proof of Lemma 6. Recall that the standard energy momentum tensor $Q_{\alpha\beta}$ on $\mathcal{M}$ is given by

$$Q_{\alpha\beta}[f] = \partial_\alpha f \partial_\beta f - \tfrac{1}{2}\mathbf{g}_{\alpha\beta}\mathbf{g}^{\mu\nu}\partial_\mu f \partial_\nu f.$$

Let $\mathscr{P}_\alpha = Q_{\alpha 0}[\phi]$. In view of (3.20), it is straightforward to derive

$$\mathbf{D}^\alpha \mathscr{P}_\alpha = \mathbf{D}^\alpha Q_{\alpha 0}[\phi] + Q_{\alpha\beta}[\phi]\mathbf{D}^\alpha \mathbf{T}^\beta = F\partial_0\phi + \tfrac{1}{2}Q_{\alpha\beta}[\phi]\pi^{\alpha\beta}.$$

Integrating in the region bounded by Σ_0, Σ_t and $\mathscr{H}$ gives

$$\begin{aligned}&\|\boldsymbol{\partial}\phi\|^2_{L^\infty_t L^2(\Sigma_t)} + \sup_{\mathscr{H}}\left(\|\nabla\phi\|^2_{L^2(\mathscr{H})} + \|L\phi\|^2_{L^2(\mathscr{H})}\right)\\ &\lesssim \|\nabla\phi_0\|^2_{L^2(\Sigma_0)} + \|\phi_1\|^2_{L^2(\Sigma_0)} + \left|\int_{\mathscr{M}} F\partial_0\phi\,\mathrm{d}\mathscr{M}\right| + \left|\int_{\mathscr{M}} Q_{\alpha\beta}[\phi]\pi^{\alpha\beta}\mathrm{d}\mathscr{M}\right| \qquad (3.22)\end{aligned}$$

where $\mathrm{d}\mathscr{M} = n\,\mathrm{d}t\,\mathrm{d}\mu_g$ denotes the volume element in the spacetime $\mathscr{M}$, with $\mathrm{d}\mu_g$ the area element in Σ_t. Using (3.14) the last term can be computed further

$$\begin{aligned}\int_{\mathscr{M}} Q_{\alpha\beta}[\phi]\pi^{\alpha\beta}\mathrm{d}\mathscr{M} &= -2\int_{\mathscr{M}} Q_{ij}[\phi]k^{ij}\mathrm{d}\mathscr{M} + \int_{\mathscr{M}} n^{-1}\nabla^i n Q_{0i}[\phi]\mathrm{d}\mathscr{M}\\ &= -2\int_{\mathscr{M}} \partial_i\phi\partial_j\phi k^{ij}\mathrm{d}\mathscr{M} + \int_{\mathscr{M}} n^{-1}\nabla^i n\partial_i\phi\partial_0\phi\,\mathrm{d}\mathscr{M}.\end{aligned}$$

One would not expect to control $\|k\|_{L^1_t L^\infty_x}$. Similar to the analysis for improving (3.7), to control the standard energy of ϕ, one has to rely on (3.9) and also use (3.11) to derive

$$\begin{aligned}\left|\int_{\mathscr{M}} Q_{\alpha\beta}[\phi]\pi^{\alpha\beta}\mathrm{d}\mathscr{M}\right| &\lesssim \|k_i.\partial^i\phi\|_{L^2(\mathscr{M})}\|\partial\phi\|_{L^2(\mathscr{M})} + \|\nabla n\|_{L^\infty(\mathscr{M})}\|\boldsymbol{\partial}\phi\|^2_{L^2(\mathscr{M})}\\ &\lesssim M^2\varepsilon\sup_{\mathscr{H}}\|\nabla\phi\|_{L^2(\mathscr{H})}\|\boldsymbol{\partial}\phi\|_{L^2(\mathscr{M})} + M\varepsilon\|\boldsymbol{\partial}\phi\|_{L^2(\mathscr{M})}\|\partial\phi\|_{L^\infty_t L^2_{\Sigma_t}}.\end{aligned}$$

Substituting the above estimate into (3.22) gives (3.21). To prove the higher order estimates, it requires more bilinear estimates than listed in Assumption 3.1. For simplicity, we skip the proof of the higher order estimates in Lemma 6. □

It follows by applying Lemma 6 to $\phi = B$ with the help of (3.19) that

$$\|\partial\boldsymbol{\partial}B\|_{L^\infty_t L^2(\Sigma_t)} \lesssim \varepsilon + M^2\varepsilon^2.$$

Using the above estimate, (3.6) and (3.17) gives

$$\|\partial A\|_{L^\infty_t L^2(\Sigma_t)} \lesssim \|\partial\mathrm{curl}\,B\|_{L^\infty_t L^2(\Sigma_t)} + \|\partial E\|_{L^\infty_t L^2(\Sigma_t)} \lesssim \varepsilon + M^2\varepsilon^2.$$

In view of the schematic formula $\partial_0 A_j = \partial_j A_0 + \mathbf{R} + \mathbf{A}\cdot\mathbf{A}$ and using (3.16) and (3.8)

$$\|\partial_0 A_j\|_{L^\infty_t L^2(\Sigma_t)} \lesssim \|\partial_j A_0\|_{L^\infty_t L^2(\Sigma_t)} + M^2\varepsilon^2 + M^4\varepsilon^3.$$

Assuming the estimates for A_0

$$\|A_0\|_{L^\infty L^4(\Sigma_t)} \lesssim M\varepsilon, \quad \|\partial_j A_0\|_{L^\infty_t L^2(\Sigma_t)} \lesssim (M\varepsilon)^{\frac{3}{2}}$$

then we arrive at

$$\|\partial A_0\|_{L^\infty_t L^2(\Sigma_t)} \lesssim \varepsilon + M^2\varepsilon^{\frac{3}{2}}.$$

By integrating the above bound in $t \in [0,1]$ with the help of (3.8), we obtain

$$\|A\|_{L^2(\Sigma_t)} \lesssim \varepsilon + M^2\varepsilon^{\frac{3}{2}},$$

as desired. We refer the reader to [42, Section 9.3] for more details.

3.3. *Construction of parametrix*

To prove the bilinear estimate (3.9) and the trilinear estimate (3.10), it relies on constructing parametrix in the curved spacetime. In [6], Klainerman and Rodnianski gave the construction of parametrix for the linear wave equation (3.20) in Lorentzian spacetime. By applying the construction in Minkowski space, they proved the standard bilinear estimates for free waves,

Proposition 7 ([6]). *Consider ϕ,ψ solutions of the flat wave equation $\Box\phi=\Box\psi=0$ and $Q(\phi,\psi)$ one of the null forms in*

$$Q_0(\phi,\psi)=\partial_\alpha\phi\cdot\partial^\alpha\psi,\quad Q_{\alpha\beta}(\phi,\psi)=\partial_\alpha\phi\cdot\partial_\beta\psi-\partial_\beta\phi\cdot\partial_\alpha\psi,\forall\alpha\neq\beta, \tag{3.23}$$

then

$$\|Q(\phi,\psi)\|_{L^2(\mathbb{R}^{3+1})}\lesssim\|\phi[0]\|_{H^2(\mathbb{R}^3)}\|\psi[0]\|_{H^1(\mathbb{R}^3)},$$

where for $f[0]=(f(0),\partial_t f(0))$,

$$\|f[0]\|_{H^\alpha(\mathbb{R}^3)}:=\|f(0)\|_{H^\alpha(\mathbb{R}^3)}+\|\partial_t f(0)\|_{H^{\alpha-1}(\mathbb{R}^3)}.$$

Moreover, they also provided the bilinear estimate for parametrix and the solution of the wave equations in the Lorentzian spacetime.

In the sequel we sketch the approach to improve the bilinear and trilinear bootstrap assumptions (3.9) and (3.10) by constructing parametrix for solution of (3.20) (see [42, Section 10]).

Let $u_\pm$ be two families of scalar functions defined on the spacetime $\mathscr{M}$ and indexed by $\omega\in\mathbb{S}^2$, satisfying the eikonal equation $\mathbf{g}^{\alpha\beta}\partial_\alpha u_\pm\partial_\beta u_\pm=0$ for each $\omega\in\mathbb{S}^2$. The initial data for $u_\pm$ are set in [43] such that $u_\pm$ at Σ_0 asymptotically approach $x\cdot\omega$. Let $\mathscr{H}^\omega_{u_\pm}$ denote the null level hypersurfaces of ${}^\omega u_\pm$. Let ${}^\omega L_\pm$ be their null normals, fixed by the condition that $\mathbf{g}({}^\omega L_\pm,\mathbf{T})=\mp1$. For any pair of functions $f_\pm$ on $\mathbb{R}^3$, define

$$\psi[f_+,f_-](t,x)=\int_{\mathbb{S}^2}\int_0^\infty \mathrm{e}^{\mathrm{i}\lambda^\omega u_+(t,x)}f_+(\lambda\omega)\lambda^2\mathrm{d}\lambda\,\mathrm{d}\omega+\int_{\mathbb{S}^2}\int_0^\infty \mathrm{e}^{\mathrm{i}\lambda^\omega u_-(t,x)}f_-(\lambda\omega)\lambda^2\mathrm{d}\lambda\,\mathrm{d}\omega.$$

We refer to the following theorem for constructing parametrix

Theorem 8 ([43, Theorem 2.11] and [45, Theorem 2.17]). *Let ϕ_0 and ϕ_1 be two scalar functions on Σ_0. Then there is a unique pair of functions (f_+,f_-) such that*

$$\psi[f_+,f_-]|_{\Sigma_0}=\phi_0,\quad \partial_t(\psi[f_+,f_-])|_{\Sigma_0}=\phi_1.$$

And $f_\pm$ satisfy the following estimates[4]*,*

$$\|\lambda^i f_\pm\|_{L^2(\mathbb{R}^3)}\lesssim\|\nabla^i\phi_0\|_{L^2(\Sigma_0)}+\|\nabla^{i-1}\phi_1\|_{L^2(\Sigma_0)},\quad i=1,2.$$

$\Box_{\mathbf{g}}\psi[f_+,f_-]$ satisfies the following estimates

$$\|\partial^i\Box_{\mathbf{g}}\psi[f_+,f_-]\|_{L^2(\mathscr{M})}\lesssim M\varepsilon(\|\nabla^{i+1}\phi_0\|_{L^2(\Sigma_0)}+\|\nabla^i\phi_1\|_{L^2(\Sigma_0)}),\quad i=0,1.$$

Due to the above theorem, associated to any pair of functions ϕ_0,ϕ_1 on Σ_0 the function $\Psi_{om}[\phi_0,\phi_1]$ defined for $(t,x)\in\mathscr{M}$

$$\Psi_{om}[\phi_0,\phi_1]=\psi[f_+,f_-].$$

Moreover, for any pair of functions $f_\pm$ on $\mathbb{R}^3$, and for any $s\in\mathbb{R}$, define the following scalar function on $\mathscr{M}$:

$$\psi_s[f_+,f_-](t,x,s)=\int_{\mathbb{S}^2}\int_0^t \mathrm{e}^{\mathrm{i}\lambda^{\omega,s}u_+(t,x)}f_+(\lambda\omega)\lambda^2\mathrm{d}\lambda\,\mathrm{d}\omega+\int_{\mathbb{S}^2}\int_0^t \mathrm{e}^{\mathrm{i}\lambda^{\omega,s}u_-(t,x)}f_-(\lambda\omega)\lambda^2\mathrm{d}\lambda\,\mathrm{d}\omega, \tag{3.24}$$

where ${}^{\omega,s}u_\pm$ are the optical functions similarly defined as for ${}^\omega u_\pm$, except that they are initialized at Σ_s as in [43]. There are analogous estimates for $\psi_s[f_+,f_-]$ as those for $\psi[f_+,f_-]$ in Theorem 8.

Next for any $s\in\mathbb{R}$, associated to $(\phi_0,\phi_1)=(0,-nF)$ a pair of functions (f_+,f_-) on Σ_s such that

$$\psi_s[f_+,f_-]|_{\Sigma_s}=0,\quad \partial_t(\psi_s[f_+,f_-]|_{\Sigma_s})=-nF.$$

Define

$$\Psi(t,s)F=\psi_s[f_+,f_-](t).$$

[4]The superscript on ∇ or ∂ denotes the order of the derivative.

Clearly

$$\Box_{\mathbf{g}}\left(\int_0^t \Psi(t,s)F(s)\mathrm{d}s\right) = F(t) + \int_0^t \Box_{\mathbf{g}}\Psi(t,s)F(s)\mathrm{d}s. \tag{3.25}$$

Similar to Theorem 8 and using (3.11), there holds

$$\|\partial^i \Box_{\mathbf{g}}\Psi(t,s)F\|_{L^2(\mathscr{M})} \lesssim M\varepsilon\|\nabla^i F\|_{L^2(\Sigma_s)}, \quad i=0,1. \tag{3.26}$$

Based on the above notations and estimates, we are ready to give the representation formula for the solution of (3.20) (see [42, Theorem 10.8]).

Proposition 9. *There holds the following representation formula for the solution of (3.20). Let*

$$\phi^{(0)} = \Psi_{om}[\phi_0,\phi_1] + \int_0^t \Psi(t,s)F^{(0)}(s,\cdot)\mathrm{d}s, \quad F^{(0)} = F$$

and for $j \geq 1$

$$\phi^{(j)} = \int_0^t \Psi(t,s)F^{(j)}(s,\cdot)\mathrm{d}s, \quad F^{(j)} = -\Box_{\mathbf{g}}\phi^{(j-1)} + F^{(j-1)}$$

then

$$\phi = \sum_{j=0}^{\infty}\phi^{(j)}.$$

Proof. We first derive from (3.25) that

$$F^{(j)} = -\int_0^t \Box_{\mathbf{g}}\Psi(t,s)F^{(j-1)}(s), \quad j \geq 2.$$

Due to (3.26), we have

$$\|\partial^i F^{(j)}\|_{L^2(\mathscr{M})} \lesssim M\varepsilon\|\nabla^i F^{(j-1)}\|_{L^2(\mathscr{M})}, \quad j \geq 2, \tag{3.27}$$

and Theorem 8 implies

$$\|\partial^i F^{(1)}\|_{L^2(\mathscr{M})} \lesssim M\varepsilon(\|\nabla^{i+1}\phi_0\|_{L^2(\Sigma_0)} + \|\nabla^i\phi_1\|_{L^2(\Sigma_0)} + \|\partial^i F\|_{L^2(M)}). \tag{3.28}$$

It follows by using Lemmas 6, (3.27), (3.28) and the construction of $\phi^{(j)}$ that

$$\begin{aligned}&\|\partial^i \boldsymbol{\partial}\phi^{(j)}\|_{L_t^\infty L^2(\Sigma_t)} + \|\partial^i F^{(j)}\|_{L^2(\mathscr{M})}\\ &\lesssim (M\varepsilon)^j(\|\nabla^{i+1}\phi_0\|_{L^2(\Sigma_0)} + \|\nabla^i\phi_1\|_{L^2(\Sigma_0)} + \|\partial^i F\|_{L^2(\mathscr{M})}), \quad i=0,1.\end{aligned} \tag{3.29}$$

Since

$$\Box_{\mathbf{g}}\left(\sum_{j=0}^{N}\phi^{(j)}\right) = F - F^{(N+1)},$$

(3.29) implies as $N \to \infty$,

$$\Box_{\mathbf{g}}\left(\sum_{j=0}^{\infty}\phi^{(j)}\right) = F.$$

Noting that

$$\phi^{(0)}|_{\Sigma_0} = \phi_0, \quad \partial_t\phi^{(0)} = \phi_1, \quad \phi^{(j)}|_{\Sigma_0} = 0, \quad \partial_t\phi^{(j)} = 0, \quad j \geq 1$$

we thus conclude the representation in Proposition 9. □

Using Proposition 9, we will prove

Proposition 10. *Denote by $\mathscr{C}(U,\nabla\phi)$ the contraction with respect to one index between a tensor U and $\nabla\phi$, for ϕ a solution of the scalar wave equation (3.20) with F,ϕ_0,ϕ_1 satisfying the estimate*

$$\|\nabla^2\phi_0\|_{L^2(\Sigma_0)} + \|\nabla\phi_1\|_{L^2(\Sigma_0)} + \|\partial F\|_{L^2(\mathscr{M})} \lesssim M\varepsilon,$$

there holds

$$\|\mathscr{C}(U,\nabla\phi)\|_{L^2(\mathscr{M})} \lesssim M\varepsilon \sup_{\omega\in\mathbb{S}^2}\|\mathscr{C}(U,{}^{\omega}\mathbf{N})\|_{L^\infty_{\omega_u}L^2(\mathscr{H}_{\omega_u})}.$$

To see the above result, we let

$$\mathfrak{C}^{\pm}[U,f]=\int_{\mathbb{S}^2}\int_0^{\infty}\mathrm{e}^{\mathrm{i}\lambda^{\omega}u_{\pm}(t,x)\omega}\mathbf{b}_{\pm}^{-1}\mathscr{C}(U,{}^{\omega}\mathbf{N}_{\pm})f(\lambda\omega)\lambda^3\mathrm{d}\lambda\mathrm{d}\omega, \tag{3.30}$$

where ${}^{\omega}\mathbf{N}_{\pm}=\nabla^{\omega}u_{\pm}/|\nabla^{\omega}u_{\pm}|$ and ${}^{\omega}\mathbf{b}_{\pm}^{-1}=|\nabla({}^{\omega}u_{\pm})|$ is the null lapse. For convenience, we may drop + or − signs in (3.30) simultaneously.

Similar to [6, Theorem 3.3], one can derive the following result (see [42, Section 11.1])

Lemma 11. *Assuming* $|{}^{\omega}\mathbf{b}^{-1}|\lesssim 1$ [5]*, there holds*

$$\|\mathfrak{C}[U,f]\|_{L^2(\mathscr{M})}\lesssim\|\lambda^2 f\|_{L^2(\mathbb{R}^3)}\sup_{\omega\in\mathbb{S}^2}\|\mathscr{C}(U,{}^{\omega}\mathbf{N})\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}.$$

Proof of Proposition 10. In view of Proposition 9, we write

$$\mathscr{C}(U,\nabla\phi)=\mathscr{C}(U,\nabla(\Psi_{om}[\phi_0,\phi_1]))+\sum_{j=0}^{\infty}\int_0^t\mathscr{C}(U,\nabla\Psi(t,s)F^{(j)}(s,\cdot))\mathrm{d}s.$$

Note

$$\mathscr{C}(U,\nabla(\Psi_{om}[\phi_0,\phi_1]))=\mathfrak{C}^{+}[U,f_+]+\mathfrak{C}^{-}[U,f_-].$$

Applying Lemma 11 to the first term in the above and also using Theorem 8 gives

$$\|\mathscr{C}(U,\nabla(\Psi_{om}[\phi_0,\phi_1]))\|_{L^2(\mathscr{M})}\lesssim\sup_{\omega\in\mathbb{S}^2}\|\mathscr{C}(U,{}^{\omega}\mathbf{N})\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}(\|\nabla^2\phi_0\|_{L^2(\Sigma_0)}+\|\nabla\phi_1\|_{L^2(\Sigma_0)}.$$

Similarly, it follows by using Theorem 8 and (3.29) that

$$\begin{aligned}&\left\|\int_0^t\mathscr{C}(U,\nabla\Psi(t,s)F^{(j)}(s,\cdot))\mathrm{d}s\right\|_{L^2(\mathscr{M})}\\&\lesssim\|\partial F^{(j)}\|_{L^2(\mathscr{M})}\sup_{\omega\in\mathbb{S}^2}\|\mathscr{C}(U,{}^{\omega}\mathbf{N})\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}\\&\lesssim(M\varepsilon)^j(\|\nabla^2\phi_0\|_{L^2(\Sigma_0)}+\|\nabla\phi_1\|_{L^2(\Sigma_0)}+\|\partial F\|_{L^2(\mathscr{M})})\sup_{\omega\in\mathbb{S}^2}\|\mathscr{C}(U,{}^{\omega}\mathbf{N})\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}.\end{aligned}$$

Summing over $j=0,\dots$, we conclude Proposition 10. □

Next, we use Propositions 10 to improve (3.9) and (3.10).

3.4. *Improvement of (3.9) and (3.10)*

We will first consider (3.9). Noting that $k_{j\cdot}=A_j$, we need control A. In view of (3.6), to improve (3.9), we first bound the leading term $\|(\mathrm{curl}\,B)^j\partial_j\phi\|_{L^2(\mathscr{M})}$.

Since B satisfies (3.18), we can apply the representation formula in Proposition 9 to B,

$$B=\Psi_{om}[\phi_0,\phi_1]+\sum_{l=0}^{\infty}\int_0^t\Psi(t,s)F^{(l)}(s,\cdot)\mathrm{d}s. \tag{3.31}$$

Note

$$(\mathrm{curl}\,B)^j\partial_j\phi=\partial_m B_n\epsilon^{jmn}\partial_j\phi,$$

where ϵ^{jmn} is the volume form on Σ_t. In this case, applying Proposition 10 with $U=\epsilon^{jmn}\partial_j\phi$, we calculate

$$\mathscr{C}(U,{}^{\omega}\mathbf{N})=\partial_j\phi\epsilon^{jmn\omega}\mathbf{N}_m=\epsilon_{AB}\nabla_A\phi,$$

and then derive by using (3.19) that

$$\begin{aligned}\|(\mathrm{curl}\,B)^j\partial_j\phi\|_{L^2(\mathscr{M})}&\lesssim(\|\nabla^2\phi_0\|_{L^2(\Sigma_t)}+\|\nabla\phi_1\|_{L^2(\Sigma_t)}+\|\partial F\|_{L^2(\mathscr{M})})\sup_{\omega\in\mathbb{S}^2}\|\mathscr{C}(U,{}^{\omega}\mathbf{N})\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}\\&\lesssim(\|\nabla^2\phi_0\|_{L^2(\Sigma_t)}+\|\nabla\phi_1\|_{L^2(\Sigma_t)}+\|\partial F\|_{L^2(\mathscr{M})})\sup_{\omega\in\mathbb{S}^2}\|\nabla\!\!\!/\phi\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}\\&\lesssim M\varepsilon\sup_{\omega\in\mathbb{S}^2}\|\nabla\!\!\!/\phi\|_{L^{\infty}_{\omega_u}L^2(\mathscr{H}_{\omega_u})}.\end{aligned}$$

[5] We assume this estimate in the sequel for simplicity. This estimate can be proved via the bootstrap argument.

Using (3.6) and (3.17), we conclude

$$\|k^j\partial_j\phi\|_{L^2(\mathscr{M})} \lesssim M\varepsilon \sup_{\omega\in\mathbb{S}^2}\|\nabla\phi\|_{L^\infty_{\omega_u}L^2(\mathscr{H}_{\omega_u})} + M^2\varepsilon^2\|\partial\phi\|_{L^\infty_t L^2(\Sigma_t)}$$

which improves (3.9).

To improve (3.10), it relies on the observation that

$$\epsilon^{jmn}\cdot{}^\omega\mathbf{N}_m Q_{j\cdots} = \mathbf{R}\cdot(\mathbf{R}\cdot L) \tag{3.32}$$

where $Q_{j\cdots}$ denotes the Bel–Robinson tensor with one component being j. Using (3.6), we derive

$$\left|\int_{\mathscr{M}} Q_{ij\gamma\delta}k^{ij}e_0^\gamma e_0^\delta\right| = \left|\int_{\mathscr{M}} Q_{ij\gamma\delta}(\operatorname{curl}B+E)^j e_0^\gamma e_0^\delta\right|. \tag{3.33}$$

Note symbolically $Q_{j\cdots} = \mathbf{R}\cdot\mathbf{R}$

$$\left|\int_{\mathscr{M}} Q_{ij\gamma\delta}\cdot Ee_0^\gamma e_0^\delta\right| \lesssim \|E\|_{L^2_tL^\infty_x}\|\mathbf{R}\|^2_{L^\infty_t L^2(\Sigma_t)} \lesssim M^4\varepsilon^4 \tag{3.34}$$

where we used (3.7) and (3.17) to derive the last inequality.

For the part of $\operatorname{curl}B$ contributed from the first term on the right-hand side of (3.31), it is reduced bound the term

$$I = \left|\int_{\mathscr{M}}\int_{\mathbb{S}^2}\int_0^\infty \mathrm{e}^{\mathrm{i}\lambda^\omega u(t,x)\omega}\mathbf{b}^{-1}\epsilon^{jmn}\cdot{}^\omega\mathbf{N}_m Q_{j\cdots} f(\lambda\omega)\lambda^3\mathrm{d}\lambda\mathrm{d}\omega\,\mathrm{d}\mathscr{M}\right|$$

where, in view of (3.19), $\|\lambda^2 f\|_{L^2(\mathbb{R}^3)} \lesssim M\varepsilon$.

Similar to improving (3.9), it is direct to bound

$$\begin{aligned}
I &\lesssim \int_{\mathbb{S}^2}\|{}^\omega\mathbf{b}^{-1}(\epsilon^{jmn}\cdot{}^\omega\mathbf{N}_m Q_{j\cdots})\int_0^\infty \mathrm{e}^{\mathrm{i}\lambda^\omega u(t,x)} f(\lambda\omega)\lambda^3\mathrm{d}\lambda\|_{L^1(\mathscr{M})}\mathrm{d}\omega\\
&\lesssim \sup_{\omega\in\mathbb{S}^2}\|{}^\omega\mathbf{b}^{-1}\|_{L^\infty(\mathscr{M})}\sup_{\omega\in\mathbb{S}^2}\|\epsilon^{jmn}\cdot{}^\omega\mathbf{N}_m Q_{j\cdots}\|_{L^2_{\omega_u}L^1(\mathscr{H}_{\omega_u})}\int_{\mathbb{S}^2}\|\lambda^3 f(\lambda\omega)\|_{L^2_\lambda}\mathrm{d}\omega\\
&\lesssim \sup_{\omega\in\mathbb{S}^2}\|\epsilon^{jmn}\cdot{}^\omega\mathbf{N}_m Q_{j\cdots}\|_{L^2_{\omega_u}L^1(\mathscr{H}_{\omega_u})}\|\lambda^2 f\|_{L^2(\mathbb{R}^3)}\\
&\lesssim M\varepsilon\sup_{\omega\in\mathbb{S}^2}\|\epsilon^{jmn}\cdot{}^\omega\mathbf{N}_m Q_{j\cdots}\|_{L^2_{\omega_u}L^1(\mathscr{H}_{\omega_u})}.
\end{aligned}$$

Substituting into (3.33) the remaining part of $\operatorname{curl}B$ contributed by the remaining term of (3.31), we can similarly obtain

$$\begin{aligned}
&\left|\int_{\mathscr{M}}\int_0^t Q_{ij\gamma\delta}(\operatorname{curl}\Psi(t,s)F^{(l)}(s,\cdot))^j e_0^\gamma e_0^\delta\mathrm{d}s\,\mathrm{d}\mathscr{M}\right|\\
&\quad\lesssim \|\partial F^{(l)}\|_{L^2(\mathscr{M})}\sup_{\omega\in\mathbb{S}^2}\|\epsilon^{jmn}\mathbf{N}_m Q_{j\cdots}\|_{L^2_{\omega_u}L^1(\mathscr{H}_{\omega_u})}.
\end{aligned}$$

Thus, with the help of (3.29) and (3.19), we arrive at

$$\left|\int_{\mathscr{M}} Q_{ij\gamma\delta}(\operatorname{curl}B)^j e_0^\gamma e_0^\delta\right| \lesssim M\varepsilon\sup_{\omega\in\mathbb{S}^2}\|\epsilon^{jmn}\mathbf{N}_m Q_{j\cdots}\|_{L^2_{\omega_u}L^1(\mathscr{H}_{\omega_u})}.$$

By using (3.7), (3.32) and (3.34), we deduce from the above that

$$\left|\int_{\mathscr{M}} Q_{ij\gamma\delta}k^{ij}e_0^\gamma e_0^\delta\right| \lesssim M\varepsilon\sup_{\omega\in\mathbb{S}^2}\|\mathbf{R}\cdot(\mathbf{R}\cdot L)\|_{L^2_{\omega_u}L^1(\mathscr{H}_{\omega_u})} + (M\varepsilon)^3 \lesssim (M\varepsilon)^3$$

which improves (3.10).

It is highly nontrivial and challenging to prove Theorem 8. The control of parametrix is established by Szeftel in [43–45]. Note that with

$$\phi_f(0,x) = \int_{\mathbb{S}^2}\int_0^\infty \mathrm{e}^{\mathrm{i}\lambda^\omega u(0,x)} f(\lambda\omega)\lambda^2\mathrm{d}\lambda\,\mathrm{d}\omega,$$

the error is

$$Ef(t,x) = \Box_{\mathbf{g}}\phi_f(t,x) = \mathrm{i}\int_{\mathbb{S}^2}\int_0^\infty \Box_{\mathbf{g}}{}^\omega u\mathrm{e}^{\mathrm{i}\lambda^\omega u(0,x)} f(\lambda\omega)\lambda^3\mathrm{d}\lambda\,\mathrm{d}\omega.$$

Note that the quantity ${}^{\omega}\mathbf{b}\Box_{\mathbf{g}}{}^{\omega}u = \mathrm{tr}\chi$ is the null expansion of $\mathscr{H}^{\omega}{}_{u}$. In Minkowski space $\Box_{\mathbf{g}}{}^{\omega}u = 0$, nevertheless $\mathrm{tr}\chi$ is nontrivial in the curved spacetime $\mathscr{M}$. Thus, controlling the null expansion is crucial for proving Theorem 8. In particular, it is necessary to derive at least the bound

$$\|\mathrm{tr}\chi\|_{L^{\infty}(\mathscr{M})} \lesssim \varepsilon$$

under the bootstrap assumptions. The control of causal geometry under regularity assumptions consistent with the bounded L^2 curvature conjecture was first given in [56–58], while an extensive set of such control is established in [44]. In the sequel, we briefly review the result of causal geometry in the rough Einstein vacuum spacetime.

4. Causal geometry of Einstein spacetime with finite curvature flux

One of the central challenges in proving the bounded L^2 curvature conjecture is controlling the causal geometry in the rough spacetime. Consistent with Theorem 3, only a bound on the curvature $\|\mathbf{R}\cdot L\|_{L^2(\mathscr{H})}$ is expected. Given this bound, Klainerman and Rodnianski provided the control of $L^{\infty}(\mathscr{H})$ bound of $\mathrm{tr}\chi$ in [56–58]. Unlike the set-up in [42], their result applies to truncated null cones rather than null hyperplanes, as used in [42] and [43–46]. The result from [56] was extended to null cones including the vertex in [59, 60], and to null cones with time foliation in [61] (a companion paper to [62]). It was also extended to the Einstein equations with certain matter fields (see [63]) and to null hyperplanes in the Einstein vacuum spacetime with canonical foliation in [64], a companion paper to [51].

Consider a null hypersurface $\mathscr{H}$ in Einstein vacuum space. Let L be the null geodesic generator of $\mathscr{H}$

$$\langle L, L\rangle = 0, \quad \mathbf{D}_L L = 0.$$

The null geodesics are parametrized by s such that $L(s) = 1$ and $s = 0$ at the initial 2-D surface S_0, which is diffeomorphic to 2-sphere.

Theorem 12 ([56]). *Consider an outgoing null hypersurface $\mathscr{H}$ intiating on a closed 2-surface S_0 diffeomorphic to $\mathbb{S}^2$, foliated by level set of s with $0 \le s \le 1$. Assume that both the set of initial data $\mathscr{I}_0$ and the curvature flux $\mathscr{R}_0 = \|\mathbf{R}\cdot L\|_{L^2(\mathscr{H})}$ are sufficiently small. Then*

$$\left\|\mathrm{tr}\chi - \frac{2}{r}\right\|_{L^{\infty}(\mathscr{H})} \lesssim \mathscr{I}_0 + \mathscr{R}_0,$$

with $r = \sqrt{(4\pi)^{-1}|S_s|}$, $|S_s|$ the area of S_s with respect to the induced metric on S_s, and additional estimates hold for $\hat{\chi}, \zeta, \underline{\chi}$.

Unlike the choice of $\underline{L}$ in Section 2, let $\underline{L}$, the conjugate null vector field, satisfy

$$\langle \underline{L}, L\rangle = -2, \langle \underline{L}, X\rangle = 0, \quad \forall X \in TS_s,$$

where S_s is the level set of s in the null hypersurface $\mathscr{H}$. Recall (2.12) for the definition of the set of connection coefficients (see also [10, Section 13.1]). Under the geodesic foliation, $\zeta + \underline{\eta} = 0$.

For $(e_a)_{a=1}^2$ the orthonormal basis on TS_s, the null components of curvature tensor $\mathbf{R}$ are given by

$$\alpha_{ab} = \mathbf{R}(L, e_a, L, e_b), \quad \beta_a = \tfrac{1}{2}\mathbf{R}(e_a, L, \underline{L}, L),$$
$$\rho = \tfrac{1}{4}\mathbf{R}(L, \underline{L}, L, \underline{L}), \quad \sigma = \tfrac{1}{4}{}^{\star}\mathbf{R}(L, \underline{L}, L, \underline{L}).$$

Below we recall a set of null structure equations for the connection coefficients and a sets of null Bianchi identities (see the detailed derivations in [10, Chapter 7]).

$$L\mathrm{tr}\chi = -\tfrac{1}{2}(\mathrm{tr}\chi)^2 - |\hat{\chi}|^2 \tag{4.1}$$

$$\nabla\!\!\!/_L\hat{\chi} = -\mathrm{tr}\chi\hat{\chi} - \alpha \tag{4.2}$$

$$\begin{aligned}
\mathrm{div}\hat{\chi} &= -\beta + \tfrac{1}{2}\nabla\mathrm{tr}\chi + \tfrac{1}{2}\mathrm{tr}\chi\zeta - \zeta\cdot\hat{\chi} && (4.3)\\
L(\mathrm{tr}\underline{\chi}) &= 2\mathrm{div}\underline{\eta} + 2\rho - \tfrac{1}{2}\mathrm{tr}\chi\cdot\mathrm{tr}\underline{\chi} - \hat{\chi}\cdot\underline{\hat{\chi}} + 2|\underline{\eta}|^2 && (4.4)\\
\nabla_L\beta + 2\mathrm{tr}\chi\beta &= \mathrm{div}\alpha + (2\zeta + \underline{\eta})\alpha && (4.5)\\
L\rho + \tfrac{3}{2}\mathrm{tr}\chi\rho &= \mathrm{div}\beta - \tfrac{1}{2}\underline{\hat{\chi}}\cdot\alpha + \zeta\cdot\beta + 2\underline{\eta}\cdot\beta && (4.6)\\
L\sigma + \tfrac{3}{2}\mathrm{tr}\chi\sigma &= -\mathrm{curl}\,\beta + \tfrac{1}{2}\underline{\hat{\chi}}\wedge\alpha - \zeta\wedge\beta - 2\underline{\eta}\wedge\beta. && (4.7)
\end{aligned}$$

Using (4.1), the $L^\infty(\mathscr{H})$ bound on $\mathrm{tr}\chi - 2/r$ can be obtained if $\int_0^s |\hat{\chi}|^2 \mathrm{d}s'$ can be bounded. Here the integral is taken along an outgoing null geodesic initiated from S_0. Since α is merely in $L^2(\mathscr{H})$, this quantity can not be controlled by using (4.2).

Using $\beta \in L^2(\mathscr{H})$, in view of the Codazzi equation (4.3), it is only expected that $\nabla\hat{\chi} \in L^2(\mathscr{H})$. Suppose the following trace inequality holds for U being $\mathrm{tr}\chi$ and $\hat{\chi}$ [6],

$$\|U\|_{L^\infty_\omega L^2_S} \lesssim \|\nabla U\|_{L^2(\mathscr{H})}. \tag{4.8}$$

Together with using (4.1), one could obtain the estimate in Theorem 12. However the trace inequality (4.8) does not hold. To have the valid bound on $\|U\|_{L^\infty_\omega L^2_S}$, one needs a stronger bound than the right-hand side of (4.8).

In [56], Klainerman and Rodnianski relied on the Besov norm and the structures of (4.3) and (4.6) to provide a sharp trace inequality to achieve the control of $\|\hat{\chi}\|_{L^\infty_\omega L^2_s(\mathscr{H})}$.

Define the following norms[7]

$$\begin{aligned}
\|F\|_{\mathscr{P}^0} &:= \sum_{\lambda\geq 1} \|P_\lambda F\|_{L^2(\mathscr{H})} + \|F\|_{L^2(\mathscr{H})}\\
\mathscr{N}_1(F) &:= \|\nabla_L F\|_{L^2(\mathscr{H})} + \|\nabla F\|_{L^2(\mathscr{H})} + \|F\|_{L^2(\mathscr{H})}\\
\|F\|_{\mathscr{B}^0} &:= \sup_{\lambda\geq 1} \|P_\lambda F\|_{L^\infty_s L^2(S_0)} + \|F\|_{L^\infty_s L^2(S_0)}.
\end{aligned}$$

Remark 13. For tensor field F, the standard Littlewood–Paley decomposition for F is applied to the components of F with respect to a set of suitably chosen parallel-transported frames. Klainerman and Rodnianski in [58] developed an intrinsic Littlewood–Paley decomposition that is directly applicable to tensor fields. For the definitions of $\mathscr{B}^0$ and $\mathscr{P}^0$, the two types of Littlewood–Paley decompositions are equivalent on a rough null hypersurface (see [59,60] for the comparison estimates between them). As we shall see when discussing the sharp trace inequality, introducing the intrinsic Littlewood–Paley decomposition is essential to completing the proof.

With $f = \int_0^s |\hat{\chi}|^2 \mathrm{d}s'$, schematically, there holds

$$\nabla_L\nabla f = 2\nabla\hat{\chi}\cdot\hat{\chi} + \cdots.$$

Hence

$$\|\nabla f\|_{\mathscr{B}^0} \lesssim \left\|\int_0^s \nabla\hat{\chi}\cdot\hat{\chi}\right\|_{\mathscr{B}^0} + \cdots.$$

Similarly, differentiating (4.1), we can derive

$$\nabla_L\nabla\mathrm{tr}\chi = -\mathrm{tr}\chi\nabla\mathrm{tr}\chi - 2\nabla\hat{\chi}\cdot\hat{\chi} + \cdots.$$

Hence

$$\|\nabla\mathrm{tr}\chi\|_{\mathscr{B}^0} \lesssim \mathscr{I}_0 + \left\|\int_0^s \nabla\hat{\chi}\cdot\hat{\chi}\right\|_{\mathscr{B}^0} + \cdots$$

[6] The norm $\|\cdot\|_{L^\infty_\omega L^2_s}$ means first taking the L^2 norm along a null geodesic initiated from S_0 along $\mathscr{H}$, followed with taking the supremum over the initial point in S_0.

[7] Along the truncated null cone $\mathscr{H}$, the area element on S_s is comparable to the one in S_0 for $0 < s \leq 1$. Therefore, we regard $L^2(S_s)$ as $L^2(S_0)$ for simplicity.

Instead of relying on $\|\nabla\hat{\chi}\|_{\mathscr{P}^0}$ to control the right-hand side, which is unobtainable with merely the bounded curvature flux, they first use (4.3) to write that

$$\nabla\hat{\chi} = \nabla\mathscr{D}_2^{-1}\beta + \nabla\mathscr{D}_2^{-1}\nabla\mathrm{tr}\chi\cdots$$

where $\mathscr{D}_2$ is the operator that takes a 2-covariant, symmetric, traceless tensor ξ into the 1-form $\mathrm{div}\xi$. Denote by $\mathscr{D}_1$ the operator that takes any 1-form ξ on S_s into the pair of functions $(\mathrm{div}\xi, \mathrm{curl}\xi)$. They observe that, in view of (4.6) and (4.7), using the normalized pair of functions $(\check{\rho},\check{\sigma})$ which are constructed to eliminate the quadratic term containing the bad term $\underline{\hat{\chi}}$ on the right-hand side,

$$\beta = \mathscr{D}_1^{-1}\{L(\check{\rho},\check{\sigma}) + \zeta\cdot\beta) + \cdots\}$$

which, symbolically, gives

$$\begin{aligned}\nabla\hat{\chi} &= \nabla\mathscr{D}_2^{-1}\mathscr{D}_1^{-1}L(\check{\rho},\check{\sigma}) + \nabla\mathscr{D}_2^{-1}\nabla\mathrm{tr}\chi + \cdots\\ &= \nabla_L\nabla\mathscr{D}_2^{-1}\mathscr{D}_1^{-1}(\check{\rho},\check{\sigma}) + \nabla\mathscr{D}_2^{-1}\nabla\mathrm{tr}\chi + [\nabla\mathscr{D}_2^{-1}\mathscr{D}_1^{-1}, L](\check{\rho},\check{\sigma}) + \cdots.\end{aligned}$$

Hence for $\nabla\hat{\chi}$, they obtain the decomposition

$$\nabla\hat{\chi} = \nabla_L P + E \tag{4.9}$$

where

$$P = \nabla\mathscr{D}_2^{-1}\mathscr{D}_1^{-1}(\check{\rho},\check{\sigma}) + \cdots, \quad E = \nabla\mathscr{D}_2^{-1}\nabla\mathrm{tr}\chi + \cdots$$

are tensors of the same type as $\nabla\hat{\chi}$.

To take advantage of such a decomposition, in [57], they establish the following sharp trace inequality with the help of integration by parts along the null geodesic.

Theorem 14 ([57, Theorem 4.1, Theorem 4.3]). *Consider the transport equation*

$$\nabla_L W = \nabla_L P\cdot F$$

with W, P, F all tensor fields tangent to S_s. There holds the sharp trace inequality

$$\|W\|_{\mathscr{B}^0} \lesssim \|W|_{S_0}\|_{B^0_{2,1}(S_0)} + \mathscr{N}_1(P)(\mathscr{N}_1(F) + \|F\|_{L^\infty_\omega L^2_s}). \tag{4.10}$$

For S_s tangent tensor fields W, E, F satisfying

$$\nabla_L W = E\cdot F$$

there holds

$$\|W\|_{\mathscr{B}^0} \lesssim \|W|_{S_0}\|_{B^0_{2,1}(S_0)} + \|E\|_{\mathscr{P}^0}(\mathscr{N}_1(F) + \|F\|_{L^\infty_\omega L^2_s}). \tag{4.11}$$

Applying Theorem 14 to $F = \hat{\chi}$ and P, E in (4.9) gives

$$\|\nabla\mathrm{tr}\chi\|_{\mathscr{B}^0} \lesssim \mathscr{I}_0 + (\mathscr{N}_1(\hat{\chi}) + \|\hat{\chi}\|_{L^\infty_\omega L^2_s})(\mathscr{N}_1(P) + \|E\|_{\mathscr{P}^0}).$$

Due to Calderon–Zygmund theorem, $\|E\|_{\mathscr{P}^0} \approx \|\nabla\mathrm{tr}\chi\|_{\mathscr{P}^0} + \cdots$, it follows by using Sobolev embedding $\|\hat{\chi}\|^2_{L^\infty_\omega L^2_s} \lesssim \|f\|_{\mathscr{B}^0}$ and Theorem 14 that

$$\|\hat{\chi}\|^2_{L^\infty_\omega L^2_s} \lesssim (\mathscr{N}_1(\hat{\chi}) + \|\hat{\chi}\|_{L^\infty_\omega L^2_s})(\mathscr{N}_1(P) + \|\nabla\mathrm{tr}\chi\|_{\mathscr{P}^0} + \cdots).$$

We then derive

$$\|\hat{\chi}\|_{L^\infty_\omega L^2_s} \lesssim \mathscr{N}_1(\hat{\chi}) + \mathscr{N}_1(P) + \|\nabla\mathrm{tr}\chi\|_{\mathscr{P}^0} + \cdots$$

Substituting the above estimate to the estimate of $\|\nabla\mathrm{tr}\chi\|_{\mathscr{B}^0}$ yields

$$\|\nabla\mathrm{tr}\chi\|_{\mathscr{B}^0} \lesssim \mathscr{I}_0 + (\mathscr{N}_1(\hat{\chi}) + \mathscr{N}_1(P) + \|\nabla\mathrm{tr}\chi\|_{\mathscr{P}^0})(\mathscr{N}_1(P) + \|\nabla\mathrm{tr}\chi\|_{\mathscr{P}^0})$$

where we dropped the additional error terms for simplicity.

$\mathscr{N}_1(\hat{\chi})$ can be bounded by using (4.2) and (4.3). $\mathscr{N}_1(P)$ can be controlled by using (4.6), (4.7), Calderon–Zygmund theorem with the help of proper commutation. Since these bounds are expected to be small, $\|\nabla\mathrm{tr}\chi\|_{\mathscr{B}^0}$ can be controlled, which gives the estimate of $|\mathrm{tr}\chi - 2/r|$.

In the above sketched proof, we neglected all the terms involved with ζ. In fact, ζ and the mass aspect function $\mu = -\mathrm{div}\zeta - \check{\rho} + |\zeta|^2$ are related in a similar pattern as the pair of quantities $(\hat{\chi}, \nabla \mathrm{tr}\chi)$. To control them, Klainerman and Rodnianski provided a trace decomposition for $\nabla\zeta$ as the one for $\nabla\hat{\chi}$. Having such a particular structure allows them to apply Theorem 14 to obtain the control of $\|\zeta\|_{L^\infty_\omega L^2_s}$ and $\|\mu\|_{\mathscr{B}^0}$, which are necessary to complete the proof for Theorem 12.

Now consider (4.10) in Theorem 14. If P is a tensor, we assume without loss of generality that P is a one form. Relative to a set of parallel-transported frames $\{X^l\}_{l=1}^m$ (see [57, Proposition 3.28]), note that

$$\|\nabla(P(X^l))\|_{L^2(\mathscr{H})} \le \|\nabla P \cdot X^l\|_{L^2(\mathscr{H})} + \|P \cdot \nabla X^l\|_{L^2(\mathscr{H})},$$

and one can obtain by transport equations that $X^l \in L^\infty$ and $\nabla X^l \in L^2_{S_0} L^\infty_s$ with bounded curvature flux. To establish the equivalence between $\|\nabla P\|_{L^2(\mathscr{H})}$ and $\sum_{l=1}^m \|\nabla(P(X^l))\|_{L^2(\mathscr{H})}$, it requires that $P \in L^\infty_{S_s} L^2_s$. With $\mathscr{D}$ either $\mathscr{D}_1$, or $\mathscr{D}_2$, from (4.9), schematically, $P \approx \mathscr{D}^{-1}(\mathbf{R} \cdot L)$. With bounded curvature flux, P would not be bounded in $L^\infty_{S_s} L^2_s$. Therefore, proving (4.10) without scalarizing the tensor field P becomes crucial. However, in the proof, one must perform a Littlewood–Paley projection, which is typically defined for scalar functions. To solve this issue, in [58], Klainerman and Rodnianski introduced a geometric Littlewood–Paley decomposition using heat flow on weakly regular 2-D closed surfaces, which can be directly applied to tensor fields. They recovered the essential analytic properties of the standard Littlewood–Paley theory for their geometric version in weakly regular 2-surfaces, such as S_s. Further simplifications for proving Theorem 14 were developed in [59,60] and [65]. In particular, a nice alternative treatment to the geometric Littlewood–Paley theory was given in [65], based on deriving improved regularity for the parallel frames.

5. The breakdown criterion of Einstein spacetime

Theorem 12 was initially motivated by the goal of proving the bounded L^2 curvature conjecture. In [7] it also provides control over the null radius of conjugacy, which played a key role in [5] in establishing the breakdown criterion for Einstein spacetimes. It is not immediately clear how controlling causal geometry under low regularity assumptions on the null cone relates to establishing a breakdown criterion for classical solutions of the Einstein vacuum equations. We will explain this connection while introducing the results of [5, 7].

Recall the standard energy argument for bounding the Bel–Robinson energy in (3.15), assuming

$$\int_{t_0}^{t_*} \{\|k\|_{L^\infty_x} + \|\nabla \log n\|_{L^\infty_x}\} \mathrm{d}t < \infty \tag{5.1}$$

the Bel–Robinson energy at Σ_t with $t \in (t_0, t_*)$ can be directly bounded in terms of its value at the initial slice Σ_0 and the bound of the quantity in (5.1).

In [5], Klainerman–Rodnianski proved that

Theorem 15 ([5]). *The Einstein vacuum spacetime with CMC foliation Σ_t with $t < 0$ can be extended beyond at $t_* < 0$ provided that*

$$\sup_{t \in [t_0, t_*)} (\|k\|_{L^\infty_x} + \|\nabla \log n\|_{L^\infty_x}) = \mathscr{K}_0 < \infty. \tag{5.2}$$

Clearly (5.2) implies (5.1) which gives the control of Bel–Robinson energy on Σ_t for $t \in (t_0, t_*)$. However only the lowest order energy is directly bounded by using (5.2) and the initial data. To extend the solution beyond t_*, they managed to bound $\|\mathbf{D}^{\le 2}\mathbf{R}\|_{L^2(\Sigma_t)}$ for $t_0 < t < t_*$ under the assumption (5.2). To control the higher order energy, they bounded $\|\mathbf{R}\|_{L^\infty}$ in the spacetime slab with the help of the equation

$$\Box_{\mathbf{g}} \mathbf{R} = \mathbf{R} \star \mathbf{R}$$

which is obtained by using Bianchi equation and (1.2).

In the same spirit to Sobolev [66] and Choquet-Bruhat [2], Klainerman and Rodnianski established a Kirchoff formula in the curved spacetime in [67] (see [68] for a simplification), by which they represented at a point p in the spacetime by

$$\mathbf{R}(p) = -\int_{\mathcal{N}^-(p,\tau)} \mathbf{A}\cdot\mathbf{R}\star\mathbf{R} + \mathscr{E} \tag{5.3}$$

where $\mathscr{E}$ denotes all other terms, $\mathbf{A}$ is a 4-covariant tensor defined as a solution of a transport equation along the backward light-cone $\mathcal{N}^-(p,\tau)$ initiated from p in the time interval $[t(p)-\tau, t(p)]$.

The representation holds within the null radius of injectivity $i_*(p)$. To define $i_*(p)$, we denote the backward null geodesic initiated from p by $\Upsilon(\omega,s)$, with $\omega\in\mathbb{S}^2$, $s(0)=p$ and $L(s)=1$ for the null geodesic generator L, normalized by $\langle L,\mathbf{T}\rangle(p)=-1$. With $\omega\in\mathbb{S}^2$ fixed, we parameterize s using the temporal parameter $\tau = t_p - t$ so that $s=s(\omega,\tau)$. $i_*(p)$ is the supremum of τ such that the exponential map sending $(\omega,\tau)\to\Upsilon(\omega,s(\tau,\omega))$ by the past null geodesics remains a global diffeomorphism between $\mathbb{S}^2\times(0,\tau)$ and its image along the backward null cone $\mathcal{N}^-(p)$. The null radius of injectivity $i_*(p)=\min(s_*(p),l_*(p))$, where $s_*(p)$ denotes the null radius of conjugacy and $l_*(p)$ is the radius of past null cut locus at p [8]. To control the higher order energies of $\mathbf{R}$, it is crucial to obtain a uniform lower bound of null radius of injectivity for all $p\in\Sigma\times(t_0,t_*)$, which is given in [7].

Next we briefly sketch the proof of [7]. We first note that only Bel-Robison energy on Σ_t, $t\in(t_0,t_*)$ and the curvature flux on $\mathcal{N}^-(p,\tau)$ with $\tau\le i_*(p)$ is bounded by universal constants. These constants depend only on the initial Bel–Robinson energy, the initial volume $|\Sigma_{t_0}|$, the initial metric bound I_0 such that $I_0^{-1}\le(g_{ij})\le I_0$, $\mathcal{K}_0$ and t_*. Consequently, only limited regularity control is obtained directly on the null cone. Prior to [7], existing results in the literature relied on pointwise bounds of the Riemann curvature tensor for controlling the injectivity radius. However, such results do not provide a uniform lower bound on the null injectivity radius in terms of universal constants.

The crucial next step is to show that

$$s_*(p) > \min(l_*(p),\delta_*)$$

where $\delta_*>0$ is a universal constant. This is achieved by showing that

$$\sup_{\mathcal{N}^-(p,\tau)}\left|\mathrm{tr}\chi - \frac{2}{s}\right| \le C \tag{5.4}$$

with $\tau=\min(l_*(p),\delta_*)$ and C a universal constant. The proof is done by rescaling the result of causal geometry analysis in [56] and [59, 60] and also by a continuity argument.

The next step is to find a good system of local space-time coordinates under which $\mathbf{g}$ is comparable with the Minkowski metric. More precisely, for a sufficiently small constant $\epsilon>0$, one needs to show that there exists a constant[9] $\delta_*>0$, depending only on ϵ and some universal constants, for which each geodesic ball $B_{\delta_*}(p)$ with $p\in\Sigma_t$ admits local coordinates $x=(x^1,x^2,x^3)$ such that under the corresponding transport coordinates $x^0=t,x^1,x^2,x^3$ the metric $\mathbf{g}=-n^2\mathrm{d}t^2+g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ with

$$|n-n(p)|\le\epsilon \quad\text{and}\quad |g_{ij}-\delta_{ij}|\le\epsilon \tag{5.5}$$

on $B_{\delta_*}(p)\times[t(p)-\delta_*,t(p)]$. Note that the first estimate can be achieved by controlling $\partial_t n$ via elliptic estimates; the second one is obtained by using (5.2) and

$$\partial_t g_{ij} = -2nk_{ij}. \tag{5.6}$$

[8] $s_*(p)$, $l_*(p)$ are both measured in terms of the temporal parameter.

[9] The constant is no greater than δ_* appeared in the above. We still use δ_* to denote it.

The existence of such local coordinates together with (5.4) will enable one to show that $\mathcal{N}^-(p,\delta_*)$ is close to the flat cone and consequently $l_*(p) > \delta_*$ [10]. Therefore the null radius of injectivity verifies

$$i_*(p) > \min(\delta_*, t(p) - t_0).$$

Then going back to (5.3), there is a uniform lower bound $\delta_* > 0$ of τ for the formula to hold. Due to the structure of $\mathbf{R}\star\mathbf{R}$, one of the curvature component can be controlled by the curvature flux along the backward null cone. Moreover one can achieve $\|\mathbf{A}\|_{L^2(\mathcal{N}^-(p,\tau))} \lesssim \tau^{1/2}$. This leads to

$$\|\mathbf{R}(t)\|_{L^\infty_\Sigma} \lesssim \tau^{1/2} \sup_{t'\in(t-\tau,t)} \|\mathbf{R}(t')\|_{L^\infty_\Sigma} + \|\mathcal{E}\|_{L^\infty_\Sigma}.$$

For $\mathcal{E}$, suppose[11]

$$|\mathcal{E}| \lesssim \sup_{t'\in[t-\tau,t-\frac{1}{2}\tau]} \tau^{-1}\|\mathbf{D}^{\leq 2}\mathbf{R}\|_{L^2(\Sigma_{t'})}.$$

Combining the above estimates implies

$$\|\mathbf{R}(t)\|_{L^\infty} \lesssim \tau^{-1} \sup_{t-2\tau\leq t'\leq t-\frac{\tau}{2}} (\|\mathbf{R}\|_{L^2(\Sigma_{t'})} + \|\mathbf{D}R\|_{L^2(\Sigma_{t'})} + \|\mathbf{D}^2\mathbf{R}\|_{L^2(\Sigma_{t'})})$$

with $\tau > 0$ sufficiently small and fixed.

Then with the step length $(1/2)\tau$, after finitely many steps, it follows from the above estimate that

$$\|\mathbf{R}\|_{H^2(\Sigma_t)} \lesssim \|\mathbf{R}\|_{H^2(\Sigma_0)}, \quad t_0 < t < t_*,$$

which enables the continuation of the solution beyond t_*.

The results on the breakdown criterion and the lower bound for the radius of injectivity of null hypersurfaces were improved in [61, 62] to depend on the weaker assumption (5.1) rather than (5.2). To control both the null radius of conjugacy and the radius of the null cut locus, the proof of Klainerman–Rodnianski relied on L^∞ bounds for k and $\nabla\log n$. With careful analysis, and assuming (5.1), it still requires universal bounds on $\int_{t(p)-\tau}^{t(p)} |k(x,t')|^2 \mathrm{d}t'$ and $\|\pi_{LL}\|_{L^\infty_\omega L^2_s(\mathcal{N}^-(p,\tau))}$ for all p in the spacetime slab to control the null radius of injectivity. In view of (5.6), bounding the first quantity yields the second estimate in (5.5). Based on a delicate bootstrap argument, [62] achieves the first bound by representing k using the wave equation for k. To obtain the second bound, it relies on decomposing $\nabla(\pi_{LL})$ into the form of $\nabla_L P + E$, followed with applying the sharp trace estimates given in Theorem 14. The latter is achieved together with controlling $\mathrm{tr}\chi - 2/s$ and other connection coefficients in [61], using the wave equation for k.

The result of Klainerman–Rodnianski in [5] was also extended by Shao in [63] to Einstein-scalar field and Einstein–Maxwell equations.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

[10] See [7] for the detailed geometric argument and [62, Section 4] for the reduction to quantitative control.

[11] Here for simplicity we skip a cutoff step, which is involved with giving the desired bound of $\mathcal{E}$.

References

[1] J. Schauder, "Das Anfangswertproblem einer quasilinearen hyperbolischen Differentialgleichung zweiter Ordnung in beliebiger Anzahl von unabhängigen Veränderlichen", *Fund. Math.* **24** (1935), no. 1, p. 213-246.

[2] Y. Choquet-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux d'érivées partielles non linéaires", *Acta Math.* **88** (1952), p. 141-225.

[3] J. Leray, *Hyperbolic Differential Equations*, The Institute for Advanced Study, Princeton, 1955.

[4] Y. Choquet-Bruhat, R. Geroch, "Global aspects of the Cauchy problem in general relativity", *Commun. Math. Phys.* **14** (1969), p. 329-335.

[5] S. Klainerman, I. Rodnianski, "On a breakdown criterion in general relativity", *J. Am. Math. Soc.* **23** (2010), no. 2, p. 345-382.

[6] S. Klainerman, I. Rodnianski, "Bilinear estimates on curved space–times", *J. Hyperbolic Differ. Equ.* **2** (2005), no. 2, p. 279-291.

[7] S. Klainerman, I. Rodnianski, "On the radius of injectivity of null hypersuraces", *J. Am. Math. Soc.* **21** (2006), p. 775-795.

[8] T. Hughes, T. Kato, J. E. Marsden, "Well-posed quasi-linear second-order hyperbolic systems with applications to nonlinear electrodynamics and general relativity", *Arch. Rational Mech. Anal.* **63** (1977), p. 273-294.

[9] A. Fischer, J. Marsden, "The Einstein evolution equations as a first-order quasi-linear symmetric hyperbolic system. I", *Commun. Math. Phys.* **28** (1972), p. 1-38.

[10] D. Christodoulou, S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series, Princeton University Press, Princeton, NJ, 1993.

[11] L. Andersson, V. Moncrief, "Elliptic-hyperbolic systems and the Einstein equations", *Ann. Henri Poincaré* **4** (2003), p. 1-34.

[12] S. Klainerman, "PDE as a unified subject", *Geom. Funct. Anal.* **1** (2000), p. 279-315.

[13] S. Klainerman, "Geometric and Fourier methods in nonlinear wave equations", in *The Set of Lectures Delivered at the IPAM Workshop in Oscillatory Integrals and PDE' March 19–25 2001, UCLA*, 2001, Available online at https://web.math.princeton.edu/~seri/papers/ucla1.pdf.

[14] H. F. Smith, "A parametrix construction for wave equations with C1, 1 coefficients", *Ann. Inst. Fourier (Grenoble)* **48** (1998), no. 3, p. 797-835.

[15] D. Tataru, "Strichartz estimates for operators with nonsmooth coefficients and the nonlinear wave equation", *Am. J. Math.* **122** (2000), p. 349-376.

[16] D. Tataru, "Strichartz estimates for second order hyperbolic operators with nonsmooth coefficients II", *Am. J. Math.* **123** (2001), p. 385-423.

[17] D. Tataru, "Strichartz estimates for second order hyperbolic operators with nonsmooth coefficients III", *J. Am. Math. Soc.* **15** (2002), p. 419-442.

[18] H. Bahouri, J. Y. Chemin, "Èquations dòndes quasilineaires et effet dispersif", *Int. Math. Res. Not.* (1999), no. 21, p. 1141-1178.

[19] H. Bahouri, J. Y. Chemin, "Eqùations dòndes quasilineaires et estimation de Strichartz", *Am. J. Math.* **121** (1999), p. 1337-1377.

[20] S. Klainerman, "A commuting vectorfield approach to Strichartz type inequalities and applications to quasilinear wave equations", *Int. Math. Res. Not.* (2001), no. 5, p. 221-274.

[21] S. Klainerman, I. Rodnianski, "Improved local well-posedness for quasilinear wave equations in dimension three", *Duke Math. J.* **117** (2003), p. 1-124.

[22] S. Klainerman, I. Rodnianski, "Rough solutions to the Einstein vacuum equations", *Ann. Math.* **161** (2005), p. 1143-1193.

[23] S. Klainerman, I. Rodnianski, "Causal structure of microlocalized rough Einstein metrics", *Ann. Math.* **161** (2005), p. 1195-1243.

[24] S. Klainerman, I. Rodnianski, "Ricci defects of microlocalized Einstein metrics", *J. Hyperbolic Differ. Equ.* **1** (2004), no. 1, p. 85-113.

[25] Q. Wang, "Rough solutions of Einstein vacuum equations in CMCSH gauges", *Commun. Math. Phys.* **328** (2014), p. 1275-1340.

[26] Q. Wang, "Causal geometry of rough Einstein CMCSH spacetime", *J. Hyperbolic Differ. Equ.* **11** (2014), no. 3, p. 563-601.

[27] Q. Wang, "A geometric approach for sharp Local well-posedness of quasilinear wave equations", *Ann. PDE* **3** (2017), article no. 12.

[28] H. F. Smith, D. Tataru, "Sharp local well-posedness results for the nonlinear wave equation", *Ann. Math.* **162** (2005), p. 291-366.

[29] B. Ettinger, H. Lindblad, "A sharp counterexample to local existence of low regularity solutions to Einstein equations in wave coordinates", *Ann. Math.* **185** (2017), no. 1, p. 311-330.

[30] H. Lindblad, "Counterexamples to local existence for quasilinear wave equations", *Math. Res. Lett.* **5** (1998), no. 5, p. 605-622.
[31] G. Ponce, T. Sideris, "Local regularity of non linear wave equations in three space dimensions", *CPDE* **18** (1993), p. 169-177.
[32] H. Lindblad, "Counterexamples to local existence for semi-linear wave equations", *Am. J. Math.* **118** (1996), no. 1, p. 1-16.
[33] S. Klainerman, M. Machedon, "Smoothing estimates for null forms and applications", *Duke Math. J.* **81** (1995), p. 99-133.
[34] S. Klainerman, M. Machedon, "On the Maxwell–Klein–Gordon equation with finite energy", *Duke Math. J.* **74** (1994), no. 1, p. 19-44.
[35] S. Klainerman, M. Machedon, "Finite energy solutions of the Yang–Mills equations in $\mathbb{R}^{3+1}$ ", *Ann. Math.* **142** (1995), no. 1, p. 39-119.
[36] S. Klainerman, M. Machedon, "Space–time estimates for null forms and the local existence theorem", *Commun. Pure Appl. Math.* **46** (1993), p. 1221-1268.
[37] J. Bourgain, "Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, part I: Schrödinger equations, part II: KdV equation", *Geom. Funct. Anal.* **3** (1993), p. 107-156, 209-262.
[38] J. Krieger, J. Lührmann, "Concentration compactness for the critical Maxwell-Klein-Gordon equation", *Ann. PDE* **1** (2015), no. 1, article no. 5.
[39] S. Oh, D. Tataru, "Global well-posedness and scattering of the (4+1)-dimensional Maxwell-Klein-Gordon equation", *Invent. Math.* **205** (2016), no. 3, p. 781-877.
[40] J. Krieger, D. Tataru, "Global well-posedness for the Yang-Mills equation in 4 + 1 dimensions: Small energy", *Ann. Math.* **3** (2017), p. 831-893.
[41] S. Oh, D. Tataru, "The threshold conjecture for the energy critical hyperbolic Yang–Mills equation", *Ann. Math.* **194** (2021), no. 2, p. 393-473.
[42] S. Klainerman, I. Rodnianski, J. Szeftel, "The bounded L2 curvature conjecture", *Invent. Math.* **202** (2015), no. 1, p. 91-216.
[43] J. Szeftel, "Parametrix for wave equations on a rough background I: regularity of the phase at initial time. II: construction and control at initial time", *Astérisque* **443** (2023), p. 1-275.
[44] J. Szeftel, "Parametrix for wave equations on a rough background III: space–time regularity of the phase", *Astérisque* **401** (2018), p. 1-338.
[45] J. Szeftel, "Parametrix for wave equations on a rough background IV: Control of the error term", *Astérisque* **444** (2023), p. 1-313.
[46] J. Szeftel, "Sharp Strichartz estimates for the wave equation on a rough background", *Ann. Sci. Éc. Norm. Supér.* **49** (2016), no. 6, p. 1279-1309.
[47] H. Lindblad, I. Rodnianski, "The weak null condition for Einstein's equations", *C. R. Acad. Sci. Paris, Ser.* **336** (2003), p. 901-906.
[48] D. Maxwell, "Rough solutions of the Einstein constraint equations", *J. Reine Angew. Math.* **590** (2006), p. 1-29.
[49] M. Dafermos, I. Rodnianski, "A new physical-space approach to decay for the wave equation with applications to black hole spacetimes", in *XVIth International Congress on Mathematical Physics 2010*, World Scientific Publishing, Hackensack, NJ, 2010, p. 421-432.
[50] Q. Wang, "Rough solutions of the 3-D compressible Euler equations", *Ann. Math.* **195** (2022), no. 2, p. 509-654.
[51] S. Czimek, O. Graf, "The spacelike-characteristic cauchy problem of general relativity in low regularity", *Ann. PDE* **8** (2022), no. 2, article no. 22.
[52] M. T. Anderson, "Cheeger-Gromov theory and applications to general relativity", in *The Einstein Equations and the Large Scale Behavior of Gravitational Fields* (P. T. Chruciel, H. Friedrich, eds.), Birkhäuser, Basel, 2004, p. 347-377.
[53] P. Petersen, "Convergence theorems in Riemannian geometry", in *Comparison Geometry (Berkeley, CA, 1993–94)*, vol. 30, Cambridge University Press, Cambridge, 1997, p. 167-202.
[54] J. Corvino, "Scalar curvature deformation and a gluing construction for the Einstein constraint equations", *Commun. Math. Phys.* **214** (2000), p. 137-189.
[55] J. Corvino, R. Schoen, "On the asymptotics for the vacuum Einstein constraint equations", *J. Differ. Geom.* **73** (2006), p. 185-217.
[56] S. Klainerman, I. Rodnianski, "Causal geometry of Einstein-vacuum spacetimes with finite curvature flux", *Invent. Math.* **159** (2005), no. 3, p. 437-529.
[57] S. Klainerman, I. Rodnianski, "Sharp trace theorems for null hypersurfaces on Einstein metrics with finite curvature flux", *Geom. Funct. Anal.* **16** (2006), p. 164-229.
[58] S. Klainerman, I. Rodnianski, "A geometric Littlewood-Paley theory", *Geom. Funct. Anal.* **16** (2006), p. 126-163.
[59] Q. Wang, *Causal geometry of Einstein vacuum spacetimes*, Phd thesis, Princeton University, 2006.

[60] Q. Wang, “On the geometry of null cones in Einstein Vacuum Spacetimes”, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26** (2009), no. 1, p. 285-328.

[61] Q. Wang, “On Ricci coefficients of null hypersurfaces with time foliation in Einstein vacuum space–time”, *Calc. Var. Partial Differ. Equ.* **46** (2013), no. 3-4, p. 461-503.

[62] Q. Wang, “Improved breakdown criterion for Einstein vacuum equation in CMC gauge”, *Commun. Pure Appl. Math.* **65** (2012), no. 1, p. 0021-0076.

[63] A. Shao, “On breakdown criteria for nonvacuum Einstein equations”, *Ann. Henri Poincaré* **12** (2011), no. 2, p. 205-277.

[64] S. Czimek, O. Graf, “The Canonical foliation on null hypersurfaces in low regularity”, *Ann. PDE* **8** (2022), no. 2, article no. 23.

[65] A. Shao, “New tensorial estimates in Besov spaces for time-dependent $(2+1)$-dimensional problems”, *J. Hyperbol. Differ. Eq.* **11** (2014), no. 4, p. 821-908.

[66] S. Sobolev, “Methodes nouvelle a resoudre le probleme de Cauchy pour les equations lineaires hyperboliques normales”, *Mat. Sb.* **1** (1936), no. 43, p. 31-79.

[67] S. Klainerman, I. Rodnianski, “A Kirchoff-Sobolev parametrix for the wave equation and applications”, *J. Hyperbolic Differ. Equ.* **4** (2007), no. 3, p. 401-433.

[68] A. Shao, “A generalized representation formula for systems of tensor wave equations”, *Commun. Math. Phys.* **306** (2011), no. 1, p. 51-82.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 379-410
https://doi.org/10.5802/crmeca.284

Review article / *Article de synthèse*

Weak cosmic censorship, trapped surfaces, and naked singularities for the Einstein vacuum equations

Censure cosmique faible, surfaces piégées et singularités nues pour les équations d'Einstein dans le vide

Yakov Shlapentokh-Rothman [a, b]

[a] Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, ON, Canada

[b] Department of Mathematical and Computational Sciences, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON, Canada

E-mail: yakovsr@math.toronto.edu

Abstract. The weak cosmic censorship conjecture posits that, generically, all singularities in General Relativity arising from regular asymptotically flat initial data should have a complete future null infinity. While this conjecture remains wide open, it has inspired many mathematical works concerning topics such as trapped surface formation and the construction of naked singularities. In this article we will review some of these works and attempt to emphasize their interconnectedness.

Résumé. La conjecture de censure cosmique faible postule que, de manière générale, toutes les singularités en relativité générale provenant de données initiales régulières asymptotiquement plates devraient avoir un infini nul futur complet. Bien que cette conjecture reste largement ouverte, elle a inspiré de nombreux travaux mathématiques portant sur des sujets tels que la formation de surfaces piégées et la construction de singularités nues. Dans cet article, nous passerons en revue certains de ces travaux et tenterons de mettre en évidence leurs interconnexions. interconnectedness.

Keywords. Weak cosmic censorship, Trapped surfaces, Naked singularities.

Mots-clés. Censure cosmique faible, Surfaces piégées, Singularités nues.

Funding. NSERC discovery grants (RGPIN-2021-02562 and DGECR-2021-00093).

Manuscript received 11 October 2024, revised and accepted 13 January 2025.

1. Introduction

We are interested here in General Relativity as a part of classical physics, and thus the Cauchy problem will be paramount for us. We now quickly recall the relevant notions. We say that the pair of a 3-dimensional Riemannian manifold (Σ, h) and a symmetric $(0,2)$-tensor k on Σ satisfy the *constraint equations* if

$$R(h) + (\mathrm{tr}_h k)^2 - |k|_h^2 = 0, \quad \mathrm{div}_h k - \nabla \mathrm{tr}_h k = 0.$$

In this case we say that (Σ, h, k) forms an *initial data set.* The fundamental theorem concerning the Cauchy problem is the following celebrated result of Choquet-Bruhat and Geroch:

Theorem 1. *[1, 2] For every sufficiently regular initial data set (Σ, h, k), there exists a unique maximal globally hyperbolic spacetime $(\mathscr{M}, g)$ solving the Einstein vacuum equations*

$$\mathrm{Ric}(g) = 0,$$

so that Σ may be embedded in $(\mathscr{M}, g)$ with first and second fundamental form h and k respectively.

The proof of Theorem 1 also yields that continuous dependence on the data holds in a suitable sense. We will assume in this article that the reader is familiar with the basics of Lorentzian geometry (such as the term "globally hyperbolic development" above in Theorem 1). See [3] for an introduction to Lorentzian geometry and causality theory.

As with any evolutionary PDE from mathematical physics, once a locally well-posed initial value problem has been established, the next natural question is, supposing we have sufficiently "nice" initial data, do we obtain a corresponding "global and singularity free" solution? Note, however, that unlike the case for the more traditional PDEs of mathematical physics, in view of the geometric nature of the Einstein equations, it is not immediately clear what "global and singularity free" should mean! Nevertheless, some necessary conditions are immediate; for example, on physical grounds it is clear that we should require that global existence at least include causal geodesic completeness. As we shall discuss in the next section, suitably understood, the famous Schwarzschild solution already illustrates the ubiquitousness of the failure of global existence for the Einstein vacuum equation.

1.1. *The Schwarzschild spacetime and geodesic incompleteness*

The first discovered non-trivial exact solution for the Einstein equations was the Schwarzschild spacetime [4]. For every $M > 0$, the *exterior region* of the Schwarzschild spacetime of mass M may be covered by so-called "Schwarzschild" coordinates $(t, r, \theta, \phi) \in \mathbb{R} \times (2M, \infty) \times \mathbb{S}^2$:

$$g_{\mathscr{S}} = -\left(1 - \frac{2M}{r}\right) \mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1} \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2).$$

It took over 30 years for the global geometry of the Schwarzschild spacetime to be properly understood. We do not have the space here to review this in full detail (see the detailed exposition in Section 2 of [5]), but it will be useful to recall some of the most salient points. The maximal extension of the Schwarzschild spacetime is a globally hyperbolic spacetime with a two-ended Cauchy hypersurface $\Sigma \approx \mathbb{R} \times \mathbb{S}^2$ and may be globally covered by a double-null coordinate system where

$$g_{\mathscr{S}} = -\Omega^2(U, V)\, \mathrm{d}U\, \mathrm{d}V + r^2(U, V)(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2),$$

for suitable functions $\Omega(U, V)$ and $r(U, V)$. Furthermore, we may rescale the U and V coordinates so that their ranges lie in a bounded set. In Figure 1 we then depict the range of the U and V coordinates with the corresponding Penrose diagram. We emphasize that each point in the diagram corresponds to a sphere in the actual spacetime.

The original Schwarzschild coordinates correspond to the rightmost diamond in the Penrose diagram. (The leftmost diamond corresponds to an isometric copy of the rightmost copy. We will mostly ignore this leftmost diamond in what follows.) Introducing a new coordinate $v = t + r + 2M\log(r - 2M)$, we obtain in (v, r, θ, ϕ) coordinates

$$g_{\mathscr{S}} = -\left(1 - \frac{2M}{r}\right) \mathrm{d}v^2 + 2\, \mathrm{d}v\, \mathrm{d}r + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2).$$

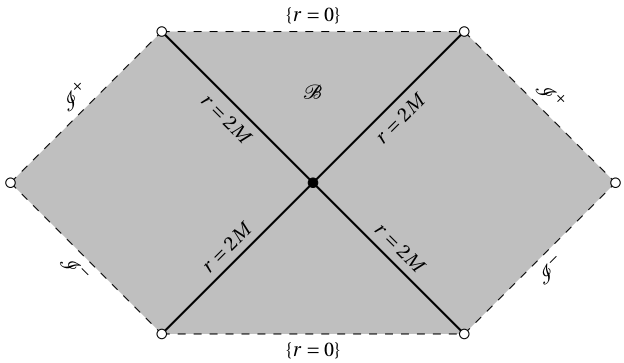


Figure 1. Penrose diagram of maximally extended Schwarzschild.

We may then let these coordinates run over the range $(v, r, \theta, \phi) \in \mathbb{R} \times (0, \infty) \times \mathbb{S}^2$, and these coordinates cover the union of the original exterior region with the region $\mathscr{B}$ from the Penrose diagram. Time-orienting by ∂_v, one now easily checks that for each $(c_0, \theta_0, \phi_0) \in \mathbb{R} \times \mathbb{S}^2$, after reparameterization, the curve

$$s \mapsto (v, r, \theta, \phi) = (c_0, -s, \theta_0, \phi_0), \quad s < 0,$$

represents a future oriented incomplete null geodesic. Furthermore, as $r \to 0$ along the geodesic, one may compute that the Kretschmann scalar $R^{\alpha\beta\gamma\delta} R_{\alpha\beta\gamma\delta}$ blows-up. It immediately follows that there is no way to acquire future null geodesic completeness by extending the Schwarzschild spacetime in a C^2 fashion[1], and thus that the Schwarzschild spacetime represents an example of singularity formation for the Cauchy problem!

Having established the existence of a singular spacetime, one immediately asks if the presence of this incompleteness is a generic feature of solutions? Could it be the case that this incompleteness is a special feature of highly symmetric spacetimes? Penrose's notion of a trapped surface, which we will discuss in the next section, quickly yields insight into these questions.

1.2. *Trapped surfaces*

Penrose's concept of a trapped surface will play a central role in this article.

Definition 1.1. *[8]* Let $\mathscr{S}$ be a closed oriented spacelike surface in a $3+1$ dimensional time-oriented Lorentzian manifold. At every point $p \in \mathscr{S}$ one may pick two linearly independent future oriented null and normal vectors L and $\underline{L}$. We say that $\mathscr{S}$ is trapped if both of the corresponding mean curvatures are negative. If one mean curvature is negative and the other vanishes, we say that $\mathscr{S}$ is marginally trapped. We say that $\mathscr{S}$ is anti-trapped if both of the corresponding mean curvatures are positive, and if one mean curvatures is positive and the other vanishes, we say that $\mathscr{S}$ is marginally anti-trapped.

Note that by the first variation formula for the area form, a surface $\mathscr{S}$ is trapped if the area form locally contracts under any flow of null vector congruences. This is physically interpreted as a sign that one is in a regime of strong gravity. A straightforward calculation shows that a sphere of radius r in the Schwarzschild spacetime is trapped if and only if $t > 0$ and $r < 2M$. This region of trapped spheres exactly corresponds to the region $\mathscr{B}$ in Figure 1.

The usefulness of the trapped surface notion follows from two fundamental facts. The first is that being a trapped surface is a manifestly stable notion. Trapped surfaces cannot disappear

[1] In fact, it is not even possible to extend maximal Schwarzschild in a C^0 fashion! See [6, 7].

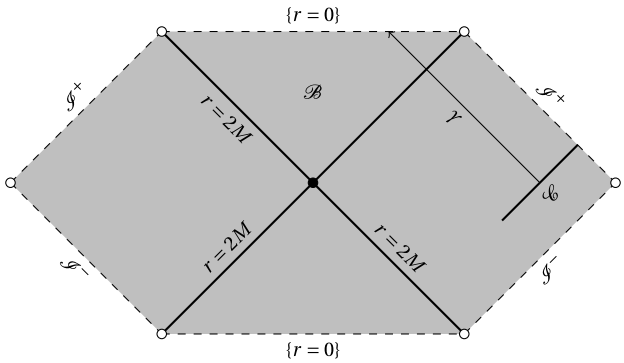


Figure 2. The hypersurface $\mathscr{C}$ and null geodesic γ in Schwarzschild from Proposition 3.

from small perturbations of a metric, and, more importantly, in view of Cauchy stability, any sufficiently small perturbation of the Cauchy data of a spacetime with trapped surfaces will still lead to a spacetime which possesses trapped surfaces. In particular, any sufficiently small perturbation of Schwarzschild's Cauchy data will still lead to a spacetime with a trapped surface. The second fact underlying the usefulness of the trapped surface notion is the following famous theorem due to Penrose.

Theorem 2. *[8] Suppose that $(\mathscr{M}, g)$ is a globally hyperbolic spacetime with a non-compact Cauchy surface and satisfies*

$$\mathrm{Ric}(V,V) \geq 0, \quad \forall V \text{ with } g(V,V) = 0.$$

Then, if $(\mathscr{M}, g)$ possesses a trapped surface, it must be future causally geodesically incomplete.

We immediately conclude that the geodesic incompleteness of Schwarzschild is a stable property to perturbations of the Cauchy data! We are thus forced to take seriously the failure of global existence in General Relativity.

1.3. *Completeness of future null infinity and the weak cosmic censorship conjecture*

While the Schwarzschild spacetime poses serious conceptual questions concerning its singularity as $r \to 0$, the spacetime geometry also suggests a "way out". On the Penrose diagram in Figure 1, we have denoted part of the boundary by $\mathscr{I}^+$. This denotes "future null infinity" and corresponds to future end points of the null hypersurfaces $t = r + 2M\log(r-2M) + c$ for any choice of $c \in \mathbb{R}$. Physically, one imagines that a far away observer propagates along a causal curve which is close to $\mathscr{I}^+$. The following proposition is straightforward to establish.

Proposition 3. *For $r_0 \gg 1$, let $\mathscr{C}$ denote the outgoing null hypersurface $\{t = r + 2M\log(r-2M)\} \cap \{r \geq r_0\}$. For each $\tilde{r} \geq r_0$ and $(\tilde{\theta}, \tilde{\phi}) \in \mathbb{S}^2$, we may consider the ingoing null geodesic $\gamma_{\tilde{r},\tilde{\theta},\tilde{\phi}}$ uniquely defined by the requirement that $\gamma_{\tilde{r},\tilde{\theta},\tilde{\phi}}(0) = \mathscr{C}|_{(r,\theta,\phi)=(\tilde{r},\tilde{\theta},\tilde{\phi})}$ and $\dot{\gamma}_{\tilde{r},\tilde{\theta},\tilde{\phi}}(0) = \partial_t - (1-(2M/r))\partial_r$. We then define $A(\tilde{r})$ to be the affine length of γ (it is immediate that A does not depend on $(\tilde{\theta}, \tilde{\phi})$). We then have that*

$$\lim_{\tilde{r}\to\infty} A(\tilde{r}) = \infty.$$

See Figure 2 for the Penrose diagram corresponding to Proposition 3. Proposition 3 has the physical interpretation that far away observers "at" future null infinity $\mathscr{I}^+$ live forever and may make observations without ever needing to interact with the Schwarzschild singularity. This good property of future null infinity is referred to as *completeness.* Our next goal is formulate a general definition of this property, but we need a few preliminaries.

Definition 4. Let (Σ, h, k) be an initial data set. We say that (Σ, h, k) is strongly asymptotically flat with n-ends if there exists a compact set K so $\Sigma \setminus K$ is diffeomorphic to n copies of $\mathbb{R}^3$ minus a ball, and there exists coordinates $\{x^a\}$ on each of these so that

$$h_{ab} = \left(1 + \frac{2M}{r}\right)\delta_{ab} + o_4(r^{-3/2}), \quad k_{ab} = o_3(r^{-5/2}),$$

where the $o_i(\cdot)$ means that up to i derivatives may be applied and that each derivative yields an extra power of decay in r.

It is a consequence of the stability of Minkowski space [9, 10], that associated to each end of a strongly asymptotically flat initial data set, the corresponding maximal development will posses an outgoing null hypersurface $\mathscr{C}$ with future complete null generators and whose geometry asymptotes to a null cone from Minkowski space. (The details of the asymptotic geometry will not be important for us here.)

We now are ready to give a definition of a complete future null infinity. We note that there are various definitions existing in the literature. Our definition is in the spirit of the definition given in [11]. A key point in both is that we do not formally define future null infinity as a boundary of a conformal compactification of spacetime. This avoids various technical regularity issues concerning such compactifications which are largely irrelevant for the physical content of the definition.

Definition 1.2. Let (Σ, h, k) be a strongly asymptotically flat initial data set and $(\mathscr{M}, g)$ be the corresponding maximal development. Associated to each asymptotically flat end, choose an outgoing null hypersurface $\mathscr{C}$ in $\mathscr{M}$ with future complete null generators whose geometry approaches that of a standard Minkowski cone. Let L denote a geodesic null normal along $\mathscr{C}$. Then let $\underline{L}$ be a choice of a transversal null vector field along $\mathscr{C}$ which is parallel transported by L. For every point $p \in \mathscr{C}$, we let $A(p)$ denote the maximal affine length of the null geodesic starting at p with initial tangent $\underline{L}$, and we let $p(s)$ denote the outgoing null geodesic which starts at p and has initial tangent vector equal to L. We then say that $(\mathscr{M}, g)$ has a *complete future null infinity* if, for every $p \in \mathscr{C}$, $\lim_{s\to\infty} A(p(s)) = \infty$.

We note that we have chosen to work with strongly asymptotically initial data sets so that we may invoke [9, 10]. However, newer proofs of the stability of Minkowski space which allow weaker versions of asymptotic flatness (cf. [12]) in principle allow for weaker asymptotic flatness assumptions in Definition 1.2. The corresponding reformulations of Definition 1.2 are straightforward to state.

Finally, we are ready for the weak cosmic censorship conjecture.

Conjecture 5 (Weak cosmic censorship). *Generic strongly asymptotically flat initial data sets possess a complete future null infinity.*

Spacetimes with an incomplete future null infinity are called *naked singularities*[2]. The earliest version of Conjecture 5 originates in [13], which, in particular, introduced the notion of a "cosmic censor" who prevents the appearance of naked singularities. However, the addition of the word "generic" into the conjecture only followed the discovery of naked singularity solutions for the spherically symmetric Einstein-scalar field system [14]. See the discussion in Section 2.3 below.

We close this section by noting some important ambiguities underlying Conjecture 5. The most apparent of these is the presence of the word "generic;" a successful resolution of Conjecture 5 will have to make this precise. A more subtle ambiguity is the fact that the maximal development of an initial data set depends on the functional framework used to solve the Einstein

[2] In this review we will restrict attention to naked singularities which arise in evolution from complete and sufficiently regular initial data. This requirement rules out solutions such as negative mass Schwarzschild or the super extremal Kerr black hole.

equations. In particular, as is the case famously with the strong cosmic censorship conjecture (see [15]), different regularity requirements on initial data or on what constitutes an admissible development of data could lead to different outcomes for Conjecture 5. See the discussions below in Sections 2.3–2.5.

1.4. *Outline of rest of the article*

Though the focus on this article is on the Einstein vacuum equations, much of our intuition about the vacuum equations derives from an influential sequence of works carried out by Christodoulou in the 1990s concerning the spherically symmetric Einstein-scalar field system. In Section 2, we will review Christodoulou's work; in particular we will discuss Christodoulou's examples of naked singularities, the study of the instabilities of naked singularities, and the central role of trapped surface formation. Finally we will briefly discuss recent work of Jaydeep Singh which yields new insights on the naked singularities of Christodouloou.

In Section 3 we will start with a discussion of Christodoulou's short-pulse method and corresponding breakthrough work on trapped surface formation for the Einstein vacuum equations. We will then survey various works which lead to significant extensions of Christodoulou's original work.

Finally, in Section 4 we will discuss the recent construction of naked singularity solutions for the Einstein vacuum equations. We will explain various notions of self-similarity and give a high level overview of the proof of existence of the naked singularities.

2. The spherically symmetric Einstein-scalar field system

Before trying to tackle the weak cosmic censorship conjecture for the Einstein vacuum equations, it is natural to work first with easier "model problems". For many PDEs of mathematical physics, such a model problem, which is consistent with asymptotic flatness, may be obtained by restricting considerations to spherically symmetric solutions. Unfortunately, for the Einstein equations, Birkhoff's theorem implies that any solution to the Einstein vacuum equations which is spherically symmetric must be static and isometric to a Schwarzschild spacetime. Arguably, the simplest way to remove this rigidity while retaining many fundamental features of the Einstein vacuum equations is to consider the spherically symmetric Einstein-scalar field system.

The unknowns for the spherically symmetric Einstein (real) scalar-field system are a $1+1$ dimensional Lorentzian manifold with boundary $(\mathcal{Q}, g)$ which arises from quotienting out the original spacetime by the spherical symmetry, an area radius function $r : \mathcal{Q} \to [0,\infty)$, and a scalar field $\phi : \mathcal{Q} \to \mathbb{R}$. The equations are

$$\begin{aligned} r\nabla_\alpha\nabla_\beta r &= \tfrac{1}{2}g_{\alpha\beta}(1-|\nabla r|_g^2) - r^2(\partial_\alpha\phi\partial_\beta\phi - \tfrac{1}{2}g_{\alpha\beta}|\nabla\phi|_g^2), \\ \nabla^\alpha(r^2\nabla_\alpha\phi) &= 0 \\ K(g) &= r^{-2}(1-|\nabla r|_g^2) + |\nabla\phi|_g^2, \end{aligned} \tag{2.1}$$

where $K(g)$ denotes the Gauss curvature of g. It is straightforward to give analogues of Definition 1.2 and Conjecture 5 for this system.

In the Sections 2.1–2.4 we review the results from an important sequence of papers due to Christodoulou [14, 16–18]. In our discussion of these papers we will follow closely the review article [11].

Finally, in Section 2.5 we discuss the recent work [19] concerning the behavior of Christodoulou's naked singularities under relatively smooth perturbations.

2.1. *Scale-invariance and BV solutions*

We set $\Gamma \doteq \partial\mathcal{Q}$ to be the set of fixed points of the $SO(3)$ action. Using that every contractible $1+1$ dimensional Lorentzian manifold is conformally equivalent to a suitable subset of $\mathbb{R}^{1+1}$, it is straightforward to show that the metric g may be written as

$$g = -\Omega^2 \, \mathrm{d}u \, \mathrm{d}v, \tag{2.2}$$

for some function $\Omega(u,v)$ and where the null coordinates u and v satisfy $v = u$ along Γ. We note that the double-null form (2.2) is preserved by any coordinate change $u \mapsto f(u)$ and $v \mapsto h(v)$ for $f', h' > 0$. The system (2.1) then becomes the following system of $1+1$ dimensional PDEs for the functions $\Omega(u,v)$, $r(u,v)$, and $\phi(u,v)$:

$$\begin{aligned}
r\frac{\partial^2 r}{\partial u \partial v} + \frac{\partial r}{\partial u}\frac{\partial r}{\partial v} &= -\frac{1}{4}\Omega^2, \\
\frac{\partial^2 r}{\partial u^2} - 2\Omega^{-1}\frac{\partial r}{\partial u}\frac{\partial \Omega}{\partial u} &= -r\left(\frac{\partial \phi}{\partial u}\right)^2, \\
\frac{\partial^2 r}{\partial v^2} - 2\Omega^{-1}\frac{\partial r}{\partial v}\frac{\partial \Omega}{\partial v} &= -r\left(\frac{\partial \phi}{\partial v}\right)^2, \\
r^2\frac{\partial^2 \log\Omega}{\partial u \partial v} - \frac{1}{4}\Omega^2 &= -r^2\frac{\partial \phi}{\partial u}\frac{\partial \phi}{\partial v} + \frac{\partial r}{\partial u}\frac{\partial r}{\partial v}, \\
r\frac{\partial^2 \phi}{\partial u \partial v} + \frac{\partial r}{\partial u}\frac{\partial \phi}{\partial v} + \frac{\partial r}{\partial v}\frac{\partial \phi}{\partial u} &= 0.
\end{aligned} \tag{2.3}$$

Closely related to the rescaling gauge freedom in the double-null coordinate system (2.2) is the fact that there is a natural group action of $\mathbb{R}_+$ on solutions to the system (2.3) defined by

$$\{r(u,v), \Omega(u,v), \phi(u,v)\} \mapsto \{a^{-1}r(au,av), \Omega(au,av), \phi(au,av)\}, \quad a \in (0,\infty). \tag{2.4}$$

Following Christoudoulou [17], we refer to solutions which are fixed points of the map (2.4) as being *scale-invariant.* In [17] Christodoulou classifies scale-invariant solutions possessing a past complete incoming null curve emanating from $(u,v) = (0,0)$. The derivative $\partial_v \phi$ will have a jump discontinuity along $\{v = 0\}$ and the spacetimes fall into three classes depending on the value of this jump

$$A \doteq \tfrac{1}{2}\left(\lim_{v \to {}^+0} r\partial_v \phi - \lim_{v \to {}^-0} r\partial_v \phi\right).$$

(1) When $A^2 < 1$, the spacetimes is global. The scalar field ϕ is constant to the past of $\{v = 0\}$ and future of $\{u = 0\}$, the transversal null derivatives of ϕ jump along $\{u = 0\}$ and $\{v = 0\}$, and in the region $\{v \geq 0\} \cap \{u \leq 0\}$ we have

$$\phi = \frac{1}{2}\log\left|\frac{u - (1+A)v}{u - (1-A)v}\right|.$$

(2) When $A^2 = 1$, the spacetime exists up to $\{u = 0\}$ where r vanishes and the scalar field ϕ blows-up.

(3) When $A^2 > 1$, the, spacetime exists up to the curve $\{u = -(A-1)v\}$ and is preceded by an *apparent horizon* along $\{u = -(A^2-1)v\}$. The spheres in between the apparent horizon and the final singular curve are all trapped.

Though these solutions are low-regularity in that the derivative of ϕ may jump, the scale-invariant spacetimes may be considered to be physically reasonable in the following sense. They admit an interpretation as an incoming spherical impulsive gravitational wave. When the amplitude of the wave is small, the wave bounces off the center and flies back out to infinity. Once the amplitude of the wave is large enough, upon collapse to the center we have singularity formation. However, generically this singularity is preceded by a trapped region which is consistent with the singularity begin hidden from infinity (cf. the region of trapped surfaces in the Schwarzschild spacetime and also the discussion in Section 2.2 below).

Christodoulou was thus inspired in [17] to build a functional framework which includes the scale-invariant solutions, has norms which are invariant under the map (2.4), and allows one to establish a local well-posedness result for the spherically symmetric Einstein-scalar field system. The class of spacetimes considered are referred to as *spacetimes of bounded variation*. In view of our space constraints in this article we will not describe in detail the norms in this space, but we note that the scalar field ϕ is required to be an absolutely continuous function, $r(\partial\phi/\partial v)$ is required to be a function of bounded variation on each hypersurface $\{u = \text{const}\}$, and $r(\partial\phi/\partial u)$ is required to be a function of bounded variation on each hypersurfae $\{v = \text{const}\}$. In addition to having scale-invariant norms defining the space, solutions of bounded variation are intimately tied to the scale-invariant solutions discussed in the paragraph above in that for any solution of bounded variation, one may consider the rescaling map (2.4) as $a \to 0$ and obtain an exactly scale invariant solution in the limit.

Finally, focusing again only on the scalar field, we note that the work [17] shows that a sufficient condition for global existence as a solution of bounded variation is, in addition to natural normalization assumptions on the geometry, for the scalar field ϕ to satisfy the following along an initial outgoing null hypersurface $\{u = \text{const}\}$:

$$\int_0^\infty |\partial_v^2(r\phi)|^2 \, \mathrm{d}v \ll 1. \tag{2.5}$$

Finally, for later use, we use the term *absolutely continuous spacetimes* to refer to the situation when the scalar field satisfies that $\partial_v(r\phi)$ and $\partial_u(r\phi)$ are absolutely continuous functions along constant u and v hypersurfaces respectively.

2.2. *Trapped surface formation*

We have seen in our earlier discussions that the Schwarzschild spacetime possesses a complete future null infinity and that all of the spheres immediately preceding the $\{r = 0\}$ singularity are trapped. This turns out not to be a coincidence and, in fact, it was shown in [16] that for the spherically symmetric Einstein-scalar field system, the existence of a single trapped surface guarantees the completeness of future null infinity. (Similar statements can in fact be made for a large class of spherically symmetric Einstein matter systems [20].) It is thus in principle quite useful to have various criteria which imply the eventual formation of a trapped surface.

Before proceeding, we define an important quantity, the *Hawking mass* m via the formula

$$1 - \frac{2m}{r} = |\nabla r|^2.$$

The Hawking mass of a sphere is always non-negative and can be considered a quasi-local measure of the energy/mass contained inside the sphere. It is useful to keep in mind that that the ratio m/r is scale-invariant in that it is left invariant under the map (2.4), and that we always have $\partial_v m \geq 0$.

Christodoulou's key result concerning trapped surface formation for the spherically symmetric Einstein-scalar field system is the following.

Theorem 6. *[16] Let $\mathcal{S}_{u_0,v_0}$ denote the sphere lying at the intersection $\{u = u_0\} \cap \{v = v_0\}$. Consider two spheres $\mathcal{S}_{0,v_0}$ and $\mathcal{S}_{0,v_1}$ lying along $\{u = 0\}$ with $v_1 > v_0 > 0$. We define the following two quantities:*

$$\delta \doteq \frac{r(\mathcal{S}_{0,v_1})}{r(\mathcal{S}_{0,v_0})} - 1, \quad \eta \doteq \frac{2(m(\mathcal{S}_{0,v_1}) - m(\mathcal{S}_{0,v_0}))}{r(\mathcal{S}_{0,v_1})}. \tag{2.6}$$

Assume that the axis Γ is regular up to the point where it intersects $\{v = v_0\}$. Then there exists a constant $c_1, c_2 > 0$ so that $0 < \delta < c_1$ and $\eta > c_2\delta\log(\delta^{-1})$ implies that there exists $\tilde{u} > 0$ with $v_0 > \tilde{u}$ and so that the sphere $\mathcal{S}_{\tilde{u},v_1}$ is a trapped sphere.

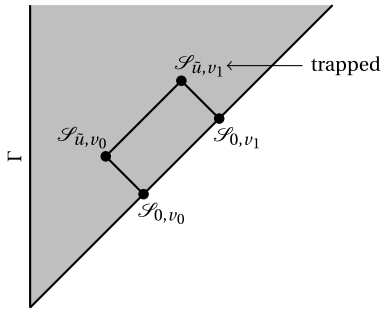


Figure 3. Penrose diagram for Theorem 6.

In Figure 3 we provide the corresponding Penrose diagram.

The quantities δ and η defined in (2.6) are both invariant under the map (2.4), and the condition that $\eta > c\delta\log(\delta^{-1})$ is sharp up to the log factor. More specifically, let $\delta > 0$ be a small constant. Then it is straightforward to construct smooth initial data for the Einstein-scalar field system along $\{u = 0\}$ so that $r|_{u=0} = v$ and so that scalar field is smooth, satisfies (2.5), is compactly supported in $v \in (1, 1+\delta)$, and also satisfies

$$\int_1^{1+\delta} |\partial_v(r\phi)|^2 \, \mathrm{d}v \gtrsim \delta, \tag{2.7}$$

where the implied constant in (2.7) depends on the implied constant in (2.5). In view of Christodoulou's theory of BV solutions (see Section 2.1 above) such data will give rise to a solution which is free of trapped surfaces. Furthermore, we will have that

$$r(\mathcal{S}_{0,1+\delta}) = 1+\delta, \quad r(\mathcal{S}_{0,1}) = 1,$$

and as consequence of the ∂_v-propagation equation for the Hawking mass[3] that

$$m(\mathcal{S}_{0,1}) = 0, \quad m(\mathcal{S}_{0,1+\delta}) \gtrsim \delta.$$

Thus, for such solutions we will have both the absence of trapped surfaces in evolution and that $\eta \gtrsim \delta$ (for η and δ as in (2.6)).

Finally, we note that it is a remarkable fact that the quantities considered in (2.6) only involve the behavior of the solution tangentially along $\{u = 0\}$ and that, other than the regularity of the axis Γ, no a priori assumptions are made about the solution on any transversal hypersurface $\{v = \text{const}\}$!

2.3. *Naked singularities*

In the work [14] Christodoulou constructed solutions to the spherically symmetric Einstein-scalar field system which correspond to naked singularities. The first step in the construction of these solutions is to introduce a generalization of the scale-invariant solutions that we discussed in Section 2.1. More specifically, we observe that the map (2.4) in fact sits inside a 1-parameter family of $\mathbb{R}_+$ actions parameterized by $k \in \mathbb{R}$ and defined by

$$\{r(u,v), \Omega(u,v), \phi(u,v)\} \mapsto \{a^{-1}r(au,av), \Omega(au,av), \phi(au,av) + k\log(a)\}, \quad a \in (0,\infty). \tag{2.8}$$

[3]This equation is

$$\partial_v m = \frac{1}{2}\left(\frac{\partial r}{\partial v}\right)^{-1}\left(1 - \frac{2m}{r}\right) r^2 (\partial_v \phi)^2.$$

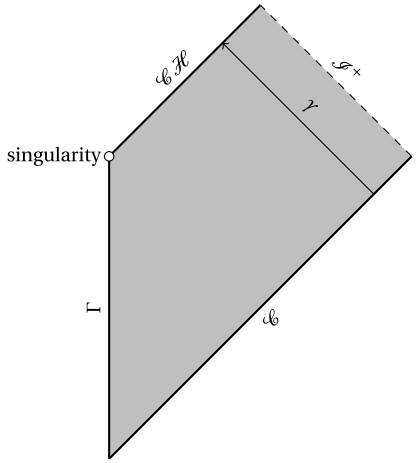


Figure 4. Penrose diagram for the naked singularities of [14] with a depicted incoming null geodesic γ of finite affine length.

A spacetime which is invariant under the map (2.8) is called *k-self similar*. Since it can be useful to not work in the double-null coordinates underlying (2.8), we observe that we can phrase k-self similarity in a coordinate invariant way: a spacetime is k-self similar if there exists a scaling vector field $\mathscr{K}$ so that

$$\mathscr{L}_{\mathscr{K}} g = 2g, \quad \mathscr{L}_{\mathscr{K}} r = r, \quad \mathscr{L}_{\mathscr{K}} \phi = -k.$$

When $k \neq 0$ it is no longer possible to write down the self-similar solutions explicitly; however, Christodoulou showed that it is still possible to reduce the k-self similar spherically symmetric Einstein-scalar system to a two dimensional autonomous system. In [14] Christodoulou carried out a systematic phase plane analysis of this system, and he found that when $0 < k^2 < 1/3$ and after a suitable truncation of the scalar field in the large r region, the k-self similar spacetimes have a Penrose diagram as depicted in Figure 4, are asymptotically flat, and moreover have an incomplete future null infinity. This showed, for the first time, that one needs the word "generic" in Conjecture 5. It will be useful for the discussion later in Section 2.4 to observe that along the past light cone of the singularity, the ratio m/r is constant and non-zero.

We make a brief note on global gauges for the k-self similar spacetimes: if one has a k-self similar spacetime in the double-null coordinate system of the map (2.4), then one may easily check that $\mathscr{K} = u\partial_u + v\partial_v$. However, it turns out that for k-self similar spacetimes, it is <u>not</u> possible to have such a coordinate system which covers the past light cone of the singularity, a hypersurface which is of fundamental importance to the analysis. Christodoulou in fact works in *Bondi coordinates* (u, r) where u is an eikonal function. Another option is to work in a rescaled double-null coordinate system where the map (2.8) and hence $\mathscr{K}$ do not take their usual form (see Appendix A of [21]).

These naked singularities do not have smooth data along the initial outgoing hypersurface $\mathscr{C}$; however, the lack of regularity there only occurs at the intersection of the past cone of the singularity and $\mathscr{C}$. At this point the tangential null derivative of the scalar field ϕ is Hölder continuous. This is more regular than initial data for a solution of bounded variation and we may thus consider the initial data to be "regular". Moreover, Christodoulou shows that the singular point is such it is impossible to extend the solution to the singular point as a solution of bounded variation. The Cauchy horizon $\mathscr{C}\mathscr{H}$ emanating from the singularity is regular in that the derivative of the scalar field extends to $\mathscr{C}\mathscr{H}$ in a Hölder continuous function; in fact, it is

possible to extend the spacetime as a k-self similar solution across $\mathscr{C}\mathscr{H}$. (Thus these spacetimes are relevant for the strong cosmic censorship conjecture.)

We will discuss the instability of these spacetimes in Section 2.4 below. Before addressing the issue of instability, a natural question that arises is whether it is possible to construct naked singularity spacetimes which do not possess any exact self-similarity. Exactly such spacetimes were constructed using a perturbative scattering approach in [21].

Though we do not have the space here to discuss these in detail, it is also interesting to construct naked singularity solutions for other spherically symmetric Einstein matter systems, for example, see [22–24]. Finally, we note that it is an open problem to rigorously construct smooth (or even C^2!) initial data for the spherically symmetric Einstein-scalar field system which leads to a naked singularity. With regards to this, it is worth observing that the physics literature around *critical collapse* (see [25, 26]) does suggest that naked singularities arising from smooth initial data should exist.

2.4. *Instability of naked singularities*

In the work [18] Christodoulou studied the instability of a general class of naked singularities for the spherically symmetric Einstein-scalar field system. Before discussing this result in its full generality, it is conceptually useful to focus on the instability of the specific (truncated) k-self similar naked singularities discussed in Section 2.3 above.

Let $\underline{\mathscr{C}}$ denote the past light cone of the singularity. The scaling vector field $\mathscr{K}$ is tangent to $\underline{\mathscr{C}}$, and we may introduce a coordinate $u < 0$ along $\underline{\mathscr{C}}$ so that $\mathscr{K}|_{\underline{\mathscr{C}}} = u\partial_u$ and so that $\underline{\mathscr{C}} \cap \mathscr{C}$ corresponds to $u = -1$. The singularity then corresponds to the limit $u \to 0$ along $\underline{\mathscr{C}}$. Let $\mathfrak{p} = (\partial_v r)^{-1}\partial_v \phi$ be a suitably normalized transversal derivative of the scalar field ϕ along $\underline{\mathscr{C}}$, so that in view of the underlying self-similarity, we have $|\mathfrak{p}| \sim u^{-1}$. Along $\underline{\mathscr{C}}$ the wave equation may be considered a transport equation for $\mathfrak{p}$:

$$\partial_u \mathfrak{p} + \frac{1+k^2}{u}\mathfrak{p} = F(u), \tag{2.9}$$

where the $1 + k^2$ and $F(u) \sim u^{-2}$ depend only on the behavior of the geometry and scalar field tangential to $\underline{\mathscr{C}}$. Even though $|\mathfrak{p}| \sim u^{-1}$, it is straightforward to see that a solution $\check{\mathfrak{p}}$ of $\mathfrak{p}$'s equation with generic initial data at $\{u = -1\}$ will have $|\check{\mathfrak{p}}| \sim |u|^{-1-k^2}$, which is a faster blow-up rate than the self-similar rate. Heuristically, we expect that if there exists a perturbation of the initial data so that the corresponding $\mathfrak{p}$ is more singular as $u \to 0$ than the self-similar rate, then we will have an instability for the naked singularity background. Optimistically, we could then hope to apply Theorem 6 to obtain the formation of a trapped surface and then deduce the existence of a complete future null infinity. We note that one may connect this instability in the evolution equation for transversal null derivatives of ϕ along $\underline{\mathscr{C}}$ to an analogous growth of the energy of null geodesics traveling along $\underline{\mathscr{C}}$; this leads us to refer to the instability as a *blue-shift instability*.

This instability can indeed be triggered by considering a perturbation of the scalar field where the solution ϕ is kept constant in the causal past of the singularity, so as to preserve the coefficients and right hand side of the Equation (2.9), and then introducing a perturbation of $\mathfrak{p}$ exactly along $\underline{\mathscr{C}} \cap \mathscr{C}$. Since the transversal derivative of ϕ thus must have a jump discontinuity, this leads to the perturbation being only in the class of solutions of bounded variation. However, it turns out that by exploiting the gap between $1+k^2$ and 1, one can also excite a similar instability with a perturbation which again keeps ϕ constant in the causal past of singularity and arranging for the perturbed $\mathfrak{p}$ to be Hölder continuous with a small k-dependent exponent $\gamma > 0$. See [19] for more details.

We now turn to the much more general instability results of [18].

Theorem 7. *[18] We parameterize initial data for the spherically symmetric Einstein-scalar field system by a choice of $\alpha = \partial_v(r\phi)$ along an initial outgoing cone $\mathscr{C}$, where $\partial_v r = 1/2$. Suppose that for some $\alpha = \alpha_0$, a function of bounded variation, the corresponding spacetime has an incomplete future null infinity and that, along the past light cone of the singularity, $(m/r) \not\to 0$ as $r \to 0$. Then there exists f_1 which is a function of bounded variation and f_2 which is an absolutely continuous function so that whenever $(c_1, c_2) \in \mathbb{R}^2 \setminus \{(0,0)\}$ the spacetime arising from the initial data $\alpha_0 + c_1 f_1 + c_2 f_2$ has a complete future null infinity.*

We note that the proof also applies to the case of a *locally naked singularities*. As in the instability of the specific naked singularities discussed above, a key role is played by a blue-shift instability along the past light cone of the singularity. However, in this more general scenario, we do not have precise information about the geometry available. Remarkably, using only the information on m/r as a starting point, Christodoulou is able to show that a blue-shift instability may be triggered by a generic (in the sense of Theorem 7) perturbation, and eventually leads to the conditions of Theorem 6 being satisfied. Finally, we note that [18] justifies the assumption on m/r along the past light cone of the singularity by a connection to continuation criteria for solutions of bounded variation from [17] (see also the corresponding discussion in [11]).

These instability proofs have been revisited with techniques developed originally for use with the Einstein vacuum equations (outside of symmetry): such techniques also allow for the study of certain instabilities generated by non-spherically symmetric and ansiostropic gravitational perturbations (see [27–29]). We note in particular the recent work [29] where An revisits the instability for the specific k-self similar naked singularitities of Christodoulou. By considering anistropic perturbations An is able to construct infinitely many more unstable directions. For these perturbations he is furthermore able to construct a corresponding anisotropic apparent horizon and to consider perturbations which vanish at the past light cone of the singularity when measured using suitable scale-invariant norms. This final fact is related to the discussion earlier in the section that the blue-shift instability can be triggered by scalar field perturbations which have a Hölder continuous first derivative as opposed to the derivative merely being absolutely continuous.

2.5. *Epilogue: the importance of the functional framework*

The naked singularities considered in Theorem 7 arose from data at least as regular as a solution of bounded variation and the perturbations considered could be as regular as a solution with $\partial_v(r\phi)$ being an absolutely continuous function. It is natural to ask what happens if we consider naked singularities arising from more regular initial data but then only allow perturbations which are also more regular. In fact, this question is already interesting if we restrict to the k-self similar naked singularities constructed by Christodoulou; for these solutions $\partial_v(r\phi)$ is Hölder continuous.

In a very interesting recent work [19], Singh has initiated the study of these questions by considering the dynamics of the wave equation $\Box_g\phi = 0$ (with no symmetry restriction) on the k-self similar naked singularity backgrounds. If the initial data for $(\partial_v r)^{-1}\partial_v\phi$ lies in a more regular Hölder space than the regularity of the background scalar field, then Singh showed that the blue-shift instability is <u>not</u> present, and the corresponding solution to $\Box_g\phi = 0$ obeys bounds which are less singular than the scalar field of the background. For initial data which lies in the same Hölder space as the background scalar field, the solution obeys bounds which are analogous to the background scalar field. In so far as solutions to the wave equation $\Box_g\phi = 0$ are a proxy for perturbations to the Einstein-scalar field system and non-spherically symmetric gravitational perturbations, these results are consistent with the k-self similar singularities being nonlinearly

stable to sufficiently regular perturbations! We conclude that the validity of the weak cosmic censorship conjecture may depend sensitively on the precise functional framework which is used both to define admissible spacetimes and "generic" perturbations. A similar connection between the validity of the strong cosmic censorship conjecture and the specific functional framework used has been an active area of research, for example, see [15, 30–32].

3. Trapped surface formation in vacuum

We have seen in Section 2 that trapped formation plays an important role in issues related to weak cosmic censorship. In moving from the spherically symmetric Einstein-scalar field system to the vacuum Einstein equations with no symmetry restrictions, the following simple sounding questions already turn out to be highly non-trivial:

Question 8. *Is it possible for a trapped surface to form in evolution from trapped, marginally trapped, anti-trapped, and marginally anti-trapped surface free initial data with one asymptotically flat end? What about the case of scattering initial data posed on past null infinity $\mathscr{I}^-$?*

Note that every Cauchy hypersurface in a Schwarzschild or a Kerr black hole spacetime is both two-ended and possesses either a trapped, anti-trapped, or marginally trapped surface. In contrast, the natural setting for the study of gravitational collapse is in the evolution of Cauchy data with one asymptotically flat end which is far from possessing any trapped or anti-trapped surface.

Question 8 was answered positively by Christodoulou in the monumental work [33]. Moreover, what is even more significant than the knowledge of a positive answer to Question 8, is that the work [33] introduced an entirely new paradigm for understanding solutions to the Einstein vacuum equations for (relatively) long times in certain large data regimes and with no symmetry. In the following sections we will start with a brief description of [33] and then we will proceed to describe various follow-up works which have introduced powerful new ideas and lead to many extensions of the original result of Christodoulou. The end result is that we currently have a quite sophisticated understanding of trapped surface formation even if we still do not quite have a theorem for the Einstein vacuum equations which is as strong as Theorem 6. (It should be noted, however, that it is far from clear that such an analogue can exist without symmetry assumptions.)

Christodoulou's work [33] and essentially all of the follow-up works on trapped surface formation have worked with the Einstein vacuum equations expressed in a double-null gauge (note, however, the interesting recent exception [34]). Many of these works have furthermore fundamentally exploited the characteristic initial value problem. We assume familiarity with these in the following sections and have included brief reviews in Sections A–C of the Appendix.

3.1. *The short pulse method of Christodoulou*

In this section we will review the short pulse method introduced in the work [33]. For a more thorough review of the original work [33] and a broader discussion of the relevant context, in addition to the original text [33], we recommend the article [35].

A key idea in [33] was to consider a certain class of characteristic initial data for the Einstein vacuum equations which allows for some "largeness" to be pumped into the system in a controlled way. We now give a brief overview of this strategy. We consider a characteristic initial value problem with data posed along two transversely intersecting null hypersurfaces $\{u = u_0\}$ and $\{v = 0\}$ for $u_0 < 0$ and $|u_0| \gg 1$. In order to eventually obtain scattering data on $\mathscr{I}^-$, one should keep in mind that we will eventually take take the limit $u_0 \to -\infty$. Along the incoming hypersurface $\{v = 0\}$ Christodoulou placed trivial Minkowskian data. Along $\{u = u_0\}$, we take $\Omega = 1$.

The key starting point for the short pulse method is to choose the remaining characteristic data along $\{u = u_0\}$ so that, in orthonormal frames on $\mathbb{S}^2$,

$$\hat{\chi}_{AB}(v,\theta^C) = |u_0|^{-1}\delta^{-1/2} f_{AB}\left(\frac{v}{\delta},\theta^C\right), \tag{3.1}$$

where $0 < \delta \ll 1$ may be taken arbitrarily small, and $f_{AB}(x,\theta^A)$ is a smooth symmetric $(0,2)$-tensor compactly supported in $(x,\theta^A) \in (0,1) \times \mathbb{S}^2$ and satisfying, for each $\theta^C \in \mathbb{S}^2$,

$$\int_0^1 (f_{AB}(x,\theta^C))^2 \,\mathrm{d}x = c_0, \quad \|f_{AB}\|_{C^k((1,2)\times\mathbb{S}^2)} \lesssim 1, \tag{3.2}$$

where $c_0 \sim 1$ is a suitable δ-independent constant and k is a suitably large δ-independent integer. The gravitational radiation field along $\mathscr{I}^-$ is given by $\lim_{u\to-\infty} \hat{\chi}_{AB}(v,\theta^C)|u|$, and, in the limit as $u_0 \to -\infty$, Equations (3.1) and (3.2) end up corresponding to a lower bound for the incoming energy per unit solid angle in each direction in the advanced time interval $[0,\delta]$.

There are now two main interrelated issues to understand. First, what is the explicit mechanism we expect to lead to the formation of the trapped surface? Second, in view of the fact that (3.1) manifestly puts us in a large data regime, why can we expect to construct enough of the spacetime to show that this mechanism takes place?

The first step in the analysis is, still at the level of the initial null hypersurface $\{u = u_0\}$, to obtain suitable estimates for all of the Ricci coefficients and null curvature components. Given the complexity of the equations of the double-null gauge (see Section A) this at first appears like a task of daunting complexity. However, Christodoulou exploits a remarkable inductive structure in the Einstein equations, which enables one to sequentially estimate all of the required terms in terms by solving transport and elliptic equations only in terms of quantities already controlled. A key point in this analysis is that while (3.1) is large in L^∞_v, it is small in any L^p_v with $p \in [1,2)$, and thus solutions to transport equations in this process often gain smallness in δ. We list here the results of this analysis for the null curvature components:

$$|\alpha| \lesssim \delta^{-3/2}|u_0|^{-1}, \quad |\beta| \lesssim \delta^{-1/2}|u_0|^{-2}, \quad |\rho| + |\sigma| \lesssim |u_0|^{-3}, \tag{3.3}$$

$$|\underline{\beta}| \lesssim \delta|u_0|^{-4}, \quad |\underline{\alpha}| \lesssim \delta^{3/2}|u_0|^{-5}. \tag{3.4}$$

We note that the u_0 weights here reflect a hierarchy consistent with *peeling* (see the discussion in [35]). It is also worth remarking that the powers of δ one sees in (3.3) and (3.4) deviate from what one would expect purely based on linear considerations.

The bulk of the technical work in [33] is concerned with showing that these hierarchies of estimates understood for the characteristic initial data can, with a few modifications, be propagated into the spacetime in a region $\{(u,v,\theta^A) \in [u_0,\check{u}] \times [0,\delta] \times \mathbb{S}^2\}$. Here $\check{u} < 0$ may be taken arbitrarily small as long δ is then taken sufficiently small depending on $\check{u}$. In order to prove this "semi-global" existence result, in addition to transport and elliptic equation estimates, Christodoulou introduces an intricate scheme of energy estimates using weighted vector fields in conjunction with the Bel–Robinson tensor. While we will not discuss these estimates in more detail here, it is worth emphasizing that it is a remarkable fact that the structure behind the initial data hierarchy can be exploited in evolution all the way from $\{u = u_0\}$ to $\{u = \check{u}\}$!

Now that we have discussed the spacetime construction, we can explain at a schematic level how the formation of the trapped surface occurs. First we must guarantee that there is no trapped surface in the initial data. In view of (3.1) and (3.2), we have that $\|\hat{\chi}\|^2_{L^\infty_{\mathbb{S}^2} L^2_{u=u_0}} \lesssim c_0|u_0|^{-2}$. It then follows by integrating Raychaudhuri's Equation (A.3), the relation $\mathscr{L}_{e_4}\mathbf{g} = \chi$, the fact that $\Omega = 1$ along $\{u = u_0\}$, and the triviality of the data along $\{v = 0\}$, that we have, along $\{u = u_0\}$

$$\left|\mathrm{tr}\chi - \frac{2}{|u_0|}\right| \lesssim \frac{c_0}{|u_0|^2}.$$

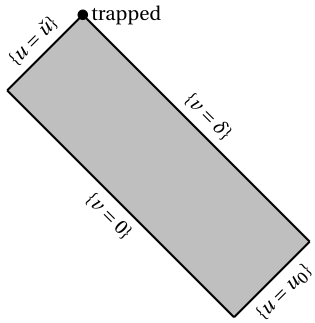


Figure 5. Penrose diagram for short pulse spacetime.

In particular, it is clear that $\mathrm{tr}\chi > 0$ and we are far from having a trapped surface along $\{u = u_0\}$. Next we would like to show that a trapped surface emerges at $(u, v) = (\check{u}, \delta)$. Our plan will be again to eventually integrate the Raychaudhuri Equation (A.3); however, first we will show $\hat{\chi}$ undergoes an amplification while evolving from $u = u_0$ to $u = \check{u}$. In order to see we use the ∇_3 propagation equation for $\hat{\chi}$ (A.5) as well as the following upper bounds which are a consequence of the estimates in the aforementioned semi-global existence result:

$$\left|\mathrm{tr}\underline{\chi} - \frac{2}{u}\right| \lesssim \delta|u|^{-2}, \quad |\Omega\underline{\omega}| \lesssim \delta|u|^{-3}, \quad |\eta| + |u|\delta^{1/2}|\nabla\eta| \lesssim \delta^{1/2}|u|^{-2},$$
$$|\hat{\underline{\chi}}| \lesssim \delta^{1/2}|u|^{-2}, \quad |\log\Omega| \lesssim \delta|u|^{-3}.$$

In particular we may write (A.5) as

$$\nabla_3\hat{\chi} - (|u|^{-1} + O(\delta|u|^{-2}))\hat{\chi} = O(|u|^{-3}).$$

Keeping in mind that δ may be assumed small depending on $\check{u}$, integrating from $\{u = u_0\}$ to $\{u = \check{u}\}$ then yields, in an orthonormal frame

$$\hat{\chi}_{AB}(v, \theta^C)|_{u=\check{u}} = |\check{u}|^{-1}\delta^{-1/2} f_{AB}\left(\frac{v}{\delta}, \theta^C\right) + O(|\check{u}|^{-2}).$$

Now we turn again to the Raychaudhuri Equation (A.3) which implies

$$\partial_v(\Omega^{-1}\mathrm{tr}\chi) \leq -|\hat{\chi}|^2. \tag{3.5}$$

Since we have Minkowskian data along $\{v = 0\}$, we integrate (3.5) from $v = 0$ along $\{u = \check{u}\}$ and obtain

$$\Omega^{-1}\mathrm{tr}\chi|_{(u,v)=(\check{u},\delta)} \leq \frac{2}{|\check{u}|} - \frac{c_0}{2|\check{u}|^2} \Rightarrow \mathrm{tr}\chi < 0.$$

In view of the smallness of $\check{u}$ we have the formation of a trapped surface.

The initial data for Christodoulou's short pulse spacetimes are given either on null hypersurfaces, or, once the limit $u_0 \to -\infty$ is taken, on past null infinity. It is natural to also ask for a version of this result with asymptotically flat initial data given along a spacelike hypersurface free of trapped surfaces. Exactly this construction was achieved in [36]. The proof used both the geometric information obtained from Christodoulou's estimates as well as gluing techniques for the constraint equations developed by Corvino–Schoen [37]. We close the section by noting the recent work [38] which provides a new proof of the formation of trapped surfaces from trapped surface free Cauchy data in the context of C^2 spacetimes. The proof is quite short compared to other such results and employs a short pulse ansatz and null gluing techniques. An interesting aspect of the proof is that the only estimates for the Einstein equations in evolution needed are those following from Cauchy stability (in particular it bypasses any need to understand a semi-global existence problem).

In Figure 5 we have drawn the Penrose diagram for Christodoulou's short pulse spacetime.

3.2. *A relaxed hierarchy and anisotropic trapped surface formation*

Soon after Christodoulou's original work [33], Klainerman–Rodnianski, focusing on the case of fixed u_0, introduced a new strategy for the semi-global existence result which leads to significant simplifications and opens the door to considering trapped surface formation by anisotropic initial data [39, 40]. The basic idea can be explain as follows (we take $u_0 = -1$ in what follows). Christodoulou's choice of data for $\hat{\chi}$ (3.1) leads to the following upper bounds for any pair of integers n and m with $n + m \leq k$ (here k is as in (3.2))

$$\delta^n \|\nabla_4^n \nabla\!\!\!/^m \hat{\chi}\|_{L^2_{\{u=-1\}}} \lesssim 1. \tag{3.6}$$

Klainerman–Rodnianski instead consider a wider class of initial data, where one only requires the following weaker bound along $u = -1$:

$$\delta^{n+m/2} \|\nabla_4^n \nabla\!\!\!/^m \hat{\chi}\|_{L^2_{\{u=-1\}}} \lesssim 1. \tag{3.7}$$

There are two main motivations for considering data of the form (3.7). First, the upper bound (3.7) is consistent with considering an initial $\hat{\chi}$ along $\{u = -1\}$ of the form

$$\hat{\chi}_{AB}(v, \theta^C) = \delta^{-1/2} f_{AB}\left(\frac{v}{\delta}, \frac{\theta^C}{\delta^{1/2}}\right).$$

This allows one to consider initial data which is highly localized in a particular angular region of the sphere. This is of interest in order to study various anisotropic phenomenon.

Before explaining the second motivation, we need to review the notion of *signature* for the Einstein vacuum equations. For any Ricci coefficient or curvature component ψ, we define

$$s(\psi) = N_4(\psi) + \tfrac{1}{2} N_A(\psi) - 1,$$

where N_4 is the number of times L appears in the definition of the component ψ and N_A is the number of angular frames:

$$s(\alpha) = 2, \quad s(\beta) = 3/2, \quad s(\rho, \sigma) = 1, \quad s(\underline{\beta}) = 1/2, \quad s(\underline{\alpha}) = 0,$$
$$s(\chi, \omega) = 1, \quad s(\zeta, \eta, \underline{\eta}) = 1/2, \quad s(\underline{\chi}, \underline{\omega}) = 0,$$
$$s(\nabla_4 \phi) = 1 + s(\phi), \quad s(\nabla\!\!\!/ \phi) = 1/2 + s(\phi), \quad s(\nabla_3 \phi) = s(\phi).$$

We also set

$$s(\phi_1 \cdot \phi_2) = s(\phi_1) + s(\phi_2),$$

where $\cdot$ stands for any type of contraction. A key point behind the usefulness of the signature concept is that the equations of the double-null gauge preserve signature in that all terms in any given equation have the same signature. Signature is furthermore preserved upon commutation with ∇_4, ∇_3, and $\nabla\!\!\!/$. The weaker bound (3.7) turns out to mesh nicely with a hierarchy of δ-bounds which is based on signature, and this ultimately allows for many simplifications in the corresponding semi-global existence result. We close by noting that this relaxed hierarchy turns to not quite be strong enough to see the amplification of $\hat{\chi}$ discussed at the end of Section 3.1, however, by interpolating the relaxed bounds with some other relatively straightforward bounds based on (3.6), the work [39] is also able to recover the trapper surface formation result of Christodoulou. The work [41] extended these methods to also allow for data on past null infinity as in Christodoulou's original work.

While the works [39, 40] allowed for highly anisotropic data (3.7) they were not quite able to prove that trapped surfaces arise from such data (though they did show the formation of so-called pre-scarred surfaces where $\mathrm{tr}\chi < 0$ on an arbitrarily large portion of a sphere). In the work [42] Klainerman, Luk, and Rodnianski were able to establish a fully anisotropic trapped surface formation result. The assumptions of the main theorem can be taken to be the same as

Christodoulou's work [33] except with one significant change, namely that instead of the left hand side of (3.2) holding for every point in $\mathbb{S}^2$, we only now require that

$$\sup_{\theta^A \in \mathbb{S}^2} \int_0^1 (f_{AB}(x,\theta^A))^2 \, \mathrm{d}x \sim 1. \tag{3.8}$$

The key idea in the work [42] was to look for a trapped surface lying along $\{v = \delta\}$ but with respect to a foliation by a new function $\tilde{u}$ as opposed to u. At a technical level, the work [42] shows that if one perturbs the null coordinate u along $\{v = \delta\}$ with the addition of a function $R(\theta^A)$ defined along $\mathbb{S}^2$, then the outgoing second fundamental forms for the new spheres have a transformation formula which involves an elliptic operator applied to the function R. It is then shown that the assumption (3.8) is sufficient to find an R satisfying a corresponding elliptic differential inequality leading to a trapped surface in the new foliation.

3.3. *Renormalization: impulsive waves and applications to trapped surface formation*

A connection can be drawn between the semi-global existence result underlying Christodoulou's trapped surface formation result and low-regularity well-posedness for the Einstein vacuum equations. Namely, since the ansatz (3.1) corresponds to the data for the metric being large in H^s, when $s > 1$, bounded when $s = 1$, and small when $s < 1$, one is lead to ask, in general, what type of initial Sobolev regularity is required for well-posedness of the initial value problem. As it turns out, a large literature exists around low-regularity well-posedness results for the Einstein equations which, in view of space constraints, we will not attempt to survey here. We will just point out that the culmination of this line of research was the remarkable collection of works [43–47] which, roughly speaking, establish that the Einstein vacuum equations are locally well-posed when the initial spacetime curvature lies in L^2 along a Cauchy hypersurface. Moreover, in view of forthcoming work of Luk–Moschidis (see Theorem 4.6 of [48]), this well-posedness result is sharp when regularity is measured in standard Sobolev spaces.

The works [49, 50] by Luk and Rodnianski initiated a new paradigm for considering certain classes of low-regularity solutions to the Einstein vacuum equations. These works were motivated by the desire to establish a local well-posedness result for the propagation (and possible interaction) of *impulsive gravitational waves* (see [51, 52]). Certain components of the curvature tensors of such spacetimes possess δ-function singularities which propagate along null hypersurfaces; for one or two propagating impulsive waves in a suitable double-null foliation, the singular curvature components correspond to α and $\underline{\alpha}$. One key insight from [49, 50], which has played a crucial role in many later developments, is that α and $\underline{\alpha}$ can be removed from the Bianchi equations for curvature in such a way that the remaining system still allows for energy estimates (see the equations of Section B). The upshot of this theory is that one has local well-posedness for characteristic initial value problems when, schematically put, the shears $\hat{\chi}$ and $\underline{\hat{\chi}}$ satisfy that

$$\sum_{i=0}^{3} \|\nabla^i \hat{\chi}\|_{L^2_{u=u_0}} + \sum_{i=0}^{3} \|\nabla^i \underline{\hat{\chi}}\|_{L^2_{v=v_0}} \doteq \mathscr{A}(\hat{\chi}, \underline{\chi}) < \infty.$$

Remarkably, Christodoulou's initial data (3.1) exactly satisfies that $\mathscr{A}(\hat{\chi}, \underline{\chi}) \lesssim 1$, and so from the "renormalized" point of view, the semi-global existence result now becomes a question of local existence. Using this, the work [50] is able to attain another proof both of Christodoulou's semi-global existence result and the corresponding trapped surface formation. Furthermore, this new proof allowed, for the first time, for non-trivial data along the incoming null hypersuface $\{v = 0\}$.

One natural question left open by all of these various results on trapped surface formation, is whether anything can be said about the short pulse spacetimes as the parameter $\delta \to 0$. This question has been addressed in [53] where the authors employ their renormalization techniques

to understand the connection between certain types of high-frequency limits of the Einstein equations and null dust solutions[4]. In particular, they show that one may take the $\delta \to 0$ (weak) limit of the short pulse spacetimes and obtain a solution to the Einstein-null dust system which also exhibits the formation of a trapped surface.

3.4. *Scale-invariant trapped surface formation*

Christodoulou's trapped surface criteria in Theorem 6 is scale invariant in that it is invariant under the rescaling map (2.4). It is natural to ask if we can also have a scale invariant criteria for trapped surface formation in vacuum, though first we must formulate the question more precisely. In any given set of local coordinates $\{x^\alpha\}$, the analogue of the rescaling map (2.4) is simply the map $x^\alpha \mapsto \lambda x^\alpha$ for $\lambda > 0$. The Sobolev space $\dot{H}^{3/2}$ is the unique homogeneous Sobolev space which is left invariant under this rescaling. Thus, we ask if one can find a trapped surface formation criteria which only requires largeness of the initial data for the metric in H^s for some $s \geq 3/2$.

The ansatz (3.1) leads to initial data which large in $H^{1+\epsilon}$, for any $\epsilon > 0$. Thus a new type of data is required. This new class of data was provided in the work [56] by An and Luk which succeeded in establishing the desired scale-invariant trapped surface formation result. While the work [56] in fact includes a large family of data which interpolates between Christodoulou like short pulse data and scale-invariant data, here we will just describe an example of the scale-invariant like data considered. Let $A \gg 1$ and $B \gg 1$ and then choose $0 < \delta \ll 1$, depending on A and B. We then take the initial data along $\{v = 0\}$ to be trivial, and suppose that $\hat{\chi}$ along $\{u = -1\}$ satisfies the following upper and lower bounds (see the discussion in the introduction to [57]):

$$\sum_{i=0}^{5} \|\nabla^i \hat{\chi}\|_{L^\infty_v L^2(\mathbb{S}^2_{-1,v})} \lesssim A, \quad \inf_{\omega \in \mathbb{S}^2} \delta^{-1} \int_0^\delta |\hat{\chi}(v,\omega)|^2 \, \mathrm{d}v \geq 4A^2. \tag{3.9}$$

Then the solution exists in a characteristic rectangle $(u, v, \theta^A) \in [-1, -\delta BA] \times [0, \delta] \times \mathbb{S}^2$, and the sphere at $(u, v) = (-\delta BA, \delta)$ is trapped.

To see the connection to $\hat{\chi}$ being small in $H^{1/2-s}$ for $s > 0$ (and hence the characteristic data being small in $H^{3/2-s}$), it is useful to first imagine that $\hat{\chi}$ is a scalar function $\tilde{\chi}(v, \theta^A)$, in which case, we could set $\tilde{\chi}(v, \theta^A) = CAf(v/\delta)$ for any smooth, non-zero, compactly supported function $f(x) : [0,1] \to \mathbb{R}$ and suitable constant C (independent of δ). This would satisfy (3.9), and we then have, for any $s \in [-1/2, 1/2]$

$$\|\tilde{\chi}\|_{H^{1/2-s}} \lesssim \|\tilde{\chi}\|_{L^2}^{1/2+s} \|\tilde{\chi}\|_{H^1}^{1/2-s} \lesssim (A\delta^{1/2})^{1/2+s} (\delta^{-1/2} A)^{1/2-s} \lesssim A\delta^s.$$

To actually construct true scale-invariant characteristic initial data for $\hat{\chi}$, one has to work a bit harder since topological obstructions prevent the existence of a nowhere vanishing trace-free $(0,2)$-tensor on $\mathbb{S}^2$. Nevertheless, it remains possible, see Appendix C of [57] for an explicit such construction.

We close the section by noting the sequence of works [57–60] where An and collaborators have studied the emergence of apparent horizons and further refined the various trapped surface formation results discussed above. We note in particular the work [58] which by use of signature considerations and a rescaling argument is able to provide new simpler proofs of the scale-invariant trapped surface formation results from [56] and also to connect the analysis to the peeling properties of the Einstein equations.

[4] These questions are motivated by Burnett's conjecture [54]; see the survey article [55].

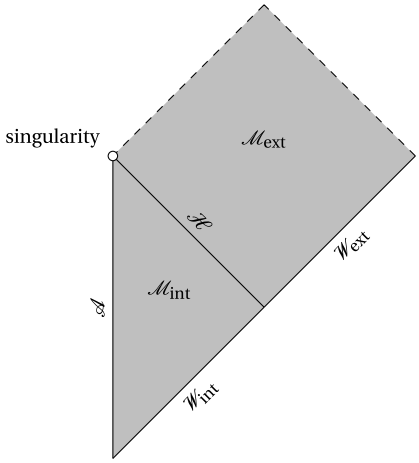


Figure 6. Penrose diagram for vacuum naked singularity.

4. Self-similarity and naked singularities in vacuum

In view of Christodoulou's construction [14] of a naked singularity for the spherically symmetric Einstein-scalar field system, it is natural to try to construct naked singularity solutions for the Einstein vacuum equations. Unfortunately, there is no direct way to import Christodoulou's construction to the Einstein vacuum equations in view of the following:

(1) Birkhoff's theorem states that all spherically symmetric solutions to the Einstein vacuum equations are isometric to Schwarzschild. Thus, we cannot expect to simply turn off the scalar field in Christodoulou's example.

(2) Christodoulou's analysis relied heavily on the fact that the spherically symmetric Einstein-scalar field system reduced, upon the imposition of self-similarity, to a two dimension autonomous system. There is no such reduction for the Einstein vacuum equations which is also consistent with asymptotic flatness.

Nevertheless, we have the following theorem:

Theorem 9. *[61, 62] There exists naked singularity solutions for the Einstein vacuum equations.*

In Figure 6 we have depicted a Penrose diagram of the naked singularity solution. (Note that other than the use of different labels for the hypersurfaces, the diagram is the same as Figure 4.) In the following sections we will describe in more details various properties of these solutions and discuss some of the main ideas which go into their construction.

4.1. *Qualitative properties and comparison with the solutions of Christodoulou*

In this section we will compare the naked singularity solutions of Christodoulou with the solutions produced by Theorem 9. When drawing analogies between solutions to the Einstein vacuum equations and solutions to the spherically symmetric Einstein-scalar field system, it is useful to identify the following:

$$\partial_u \phi \leftrightarrow \hat{\underline{\chi}}, \quad \partial_v \phi \leftrightarrow \hat{\chi}.$$

We will see that both classes of naked singularities share many qualitative properties.

As with Christodoulou's naked singularities, the vacuum naked singularities arise from asymptotically flat data along a null hypersurface. We recall that for Christodoulou's solutions, the initial

characteristic data for the scalar field satisfies that $(\partial_v r)^{-1}\partial_v\phi$ is smooth except when it crosses the past light cone of the singular point where it is Hölder continuous. The initial characteristic data for the solutions of Theorem 9 can be considered analogous in that, with respect to a suitable v-coordinate and for any $N \in \mathbb{Z}$ with $N \gg 1$, Theorem 9 produces solutions so that the initial $\hat{\chi}$ is C^N except where it crosses the past light cone of the singularity. Across the past light cone of the singularity, we have that $\hat{\chi}$ is Hölder continuous. The loss of regularity for $\hat{\chi}$ is only with respect to ∂_v; it in fact is $C^N_{\mathbb{S}^2}C^{1,\gamma}_v$.

We cannot appeal to Christodoulou's well-posedness result in BV in order to justify that this class of initial data is "sufficiently regular". Instead we start by observing that $\hat{\chi}$ being Hölder continuous is more than what it is needed for the local theory developed by [49, 50]. The analysis from [49, 50] did not attempt to cover the solution near the "axis" but, for this type of data with a Hölder continuous $\hat{\chi}$, one expects such a well-posedness result to hold.

We now turn to a discussion of the main singularity of the spacetime. We have already mentioned that Christodoulou's naked singularities cannot be extended to the singular point in such a way as to remain a solution of bounded variation. One way to see this is to observe that along the past light cone of singularity $\mathscr{C}$ we have

$$\int_{\mathscr{C}} |\partial_u\phi|\,\mathrm{d}u = \infty.$$

For the naked singularities of Theorem 9 we have the following analogous fact. There exists (at least one) null geodesic γ converging to the singularity so that

$$\int_{\gamma} |\hat{\chi}|\,\mathrm{d}s = \infty.$$

In fact, one can also check that there exists Jacobi fields which blow-up along γ and that the singularity is a *strong singularity* in the sense of Tipler [63].

4.2. *Self-similarity for the Einstein vacuum equations*

A fundamental role in the naked singularity construction is played by various types of *self-similarity*. In Section 4.2.1 we discuss an important precursor, the self-similar spacetimes of Fefferman–Graham. Then in Section 4.2.2 we will discuss a new type of self-similar spacetimes and discuss why one could expect them to be useful for the construction of naked singularities.

4.2.1. *The self-similarity of Fefferman–Graham*

Fefferman–Graham [64, 65] were the first to systematically study a class of self-similar solutions to the Einstein vacuum equations without any symmetry assumptions. Their class of spacetimes was the following:

Definition 10. We say that a spacetime $(\mathscr{M}, g)$ is Fefferman–Graham self-similar (FGSS) if $(\mathscr{M}, g)$ may be equipped with a double-null foliation (A.1) which covers $u \in (-\infty, 0)$ and at least one of $(v/-u) \in [0, c)$ or $(v/-u) \in (-c, 0]$ for some small constant c, and so that $K = u\partial_u + v\partial_v$ satisfies

$$\mathscr{L}_K g = 2g. \tag{4.1}$$

Using (4.1), a straightforward computation shows that the metric components Ω, b, and $\mathbf{g}$ of a FGSS spacetime must satisfy in coordinate frames that

$$\begin{gathered}\Omega(u,v,\theta^A) = \tilde{\Omega}\left(\frac{v}{u},\theta^A\right), \quad b^A(u,v,\theta^B) = u^{-1}\tilde{b}^A\left(\frac{v}{u},\theta^B\right),\\ \mathbf{g}_{AB}(u,v,\theta^C) = u^2\tilde{\mathbf{g}}_{AB}\left(\frac{v}{u},\theta^C\right),\end{gathered} \tag{4.2}$$

for some $\tilde{\Omega}$, $\tilde{b}$, and $\tilde{\mathbf{g}}$.

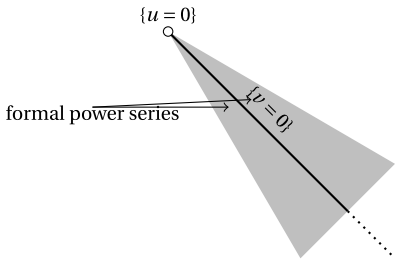


Figure 7. Region covered the Fefferman–Graham formal expansions.

Fefferman–Graham did not actually work in the double-null gauge, but it is possible to translate their study of self-similar solutions to the double-null gauge. As a preliminary step in their program to understand local conformal invariants, Fefferman–Graham studied formal power series expansions for FGSS spacetimes. If we specialize their results concerning these formal expansions to spacetime dimension 4 and recast them in the language of the double null-gauge , we can informally[5] phrase their result as follows:

Theorem 11. *For every smooth metric $\mathbf{g}_0$ on $\mathbb{S}^2$ and smooth $\mathbf{g}_0$-trace-free symmetric* $(0,2)$*-tensor k_0 on $\mathbb{S}^2$, there exists a smooth FGSS metric g defined in the double-null gauge (A.1) for $\{0 \leq (v/-u) < 1/2\} \cap \{u \in (-\infty, 0)\}$ so that g has a unique formal Taylor series expansion in $v/-u$, $g \approx \sum_{j=1}^{\infty} g_j$, and*

(1) *The induced metric $\mathbf{g}$ on $\mathbb{S}^2_{u,v}$ satisfies $\mathbf{g}|_{v=0} = u^2 \mathbf{g}_0$ in the coordinate frame, the quantities b and* $\log\Omega$ *associated to g vanish along $\{v = 0\}$, and the restriction of $\hat{\chi}_{AB}$ to $\{v = 0\}$ satisfies $\hat{\chi}_{AB} = u\,(k_0)_{AB}$ in the coordinate frame.*

(2) *For every sufficiently large positive integer n we have* $\mathrm{Ric}(\sum_{j=1}^{n} g^{(j)})(X,Y) = O(|v/u|^n)$*, for $X, Y \in \{\partial_v, \partial_u, u^{-1}\partial_{\theta^A}\}$.*

In Figure 7 we depict the region of spacetime formally covered by these expansions:

We note that if $\mathbf{g}_0$ and k_0 are analytic, then the formal expansions produced by the above theorem will actually converge (see [65]). If we leave the analytic category, then we can appeal to the work [66] to show the existence of a true solution in the region to the future of $\{v = 0\}$ provided $\mathbf{g}$ and k_0 lie in suitable Sobolev spaces.

In particular, in view of (4.2) it is plausible that a FGSS spacetime cannot be extended in a regular fashion to $u = 0$ along curves of constant v/u. Thus it may appear a reasonable strategy to use a FGSS spacetime to try to build a naked singularity[6]. However, the formal theory of Fefferman–Graham and its extensions in [66] only concern the description of the spacetime locally near $\{v = 0\}$, while the construction of a naked singularity involves a global analysis, that is, we need our spacetime to contain a hypersurface corresponding to complete and asymptotically flat initial data. The global analysis turns out to be problematic because there is no known way to extend a FGSS spacetime to the past of $\{v = 0\}$ in such a way which is compatible with the existence of a past complete Cauchy hypersurface other than by gluing a *flat* spacetime to the past of $\{v = 0\}$. Unfortunately, such a flat extension to the past of $\{v = 0\}$ prevents the resulting spacetime from being considered (part of) a naked singularity for (at least) the following two reasons:

[5] See [66, 67] for precise statements.

[6] We note that even if one did not know about the work of Fefferman–Graham, in view of the long tradition of the use of self-similarity to produce singular solutions to PDEs in mathematical physics, it would be reasonable to try to search for self-similar solutions of the Einstein equations as part of the pursuit of naked singularities.

(1) If the spacetime is flat to the past of $\{v = 0\}$, then it is a consequence of the formal theory of Fefferman–Graham that in order for the spacetime to not be globally flat, we must allow $\hat{\chi}$ to have a jump discontinuity along $\{v = 0\}$. This lowers the regularity of whatever we pick to be our Cauchy hypersurface, the resulting spacetime represents an impulsive gravitational wave with an ingoing spherical front, and it is difficult to argue that the "singularity" at $\{u = 0\}$ represents a loss of regularity.
(2) If the spacetime is flat to the past of $\{v = 0\}$ then any causal curve entering the "singularity" at $\{u = 0\}$ will only pass through a flat spacetime and thus not actually encounter any truly singular behavior.

4.2.2. *Twisted self-similarity*

One of the key ideas in our work [61] was the introduction of a generalization of FGSS spacetimes which is consistent with the possibility of non-flat, past complete extensions to the past of $\{v = 0\}$. We will give a brief introduction in this section to this new type of self-similarity. We direct the reader to the work [67] for a systematic treatment.

The rigidity of FGSS spacetimes that we have mentioned in the previous section has its source in the following three closely related facts about the geometry of the null hypersurface $\{v = 0\}$ in the FGSS spacetimes:

(1) The induced (indefinite) metric along $\{v = 0\}$ is $u^2 \tilde{\not{g}}_{AB}(\theta^C)\, d\theta^A\, d\theta^B$. In view of the uniformization theorem, we can write $\tilde{\not{g}}_{AB} = e^{2\varphi} \mathring{\not{g}}_{AB}$ where $\mathring{\not{g}}_{AB}$ denotes the round metric on the sphere and φ is a function on $\mathbb{S}^2$. Then, defining a new coordinate $\tilde{u} \doteq u e^{\varphi}$, in the $(\tilde{u}, \theta^A)$ coordinates, the induced metric along $\{v = 0\}$ becomes $\tilde{u}^2 \mathring{\not{g}}_{AB}\, d\theta^A\, d\theta^B$. This is identical to the induced metric along $\{v = 0\}$ in Minkowski space!
(2) A computation shows that all components of the curvature tensor which are tangential to $\{v = 0\}$ must vanish.
(3) The orbits of the restriction of the conformal Killing field $K = u\partial_u + v\partial_v$ to $\{v = 0\}$ coincide with the orbits of the null generators of $\{v = 0\}$. Equivalently, the shift vector vanishes on $\{v = 0\}$.

In [61] we modified Definition 10 by only requiring that the double-null coordinate system was valid for $(v/-u) \in (0, c)$ instead of $[0, c)$. We then required that the spacetime extends to $\hat{v} = 0$ in a new coordinate system $(\hat{v}, u, \theta^A)$ where

$$\hat{v} \doteq v^{1-2\kappa}$$

for some $0 < \kappa \ll 1$. We refer to such a self-similar spacetime as a κ-self-similar (κSS) spacetime. Of fundamental important for the construction is the ingoing Raychaudhuri Equation (A.4) along the null hypersurface $\{\hat{v} = 0\}$. If $\lim_{v\to 0} \Omega^2 (v/-u)^{2\kappa} = 1$, then this becomes the following equation along $\{\hat{v} = 0\}$:

$$\mathrm{div}\, b - \mathscr{L}_b \mathrm{div}\, b - \tfrac{1}{2}(\mathrm{div}\, b)^2 = \tfrac{1}{4}|\nabla\!\!\!/ \hat{\otimes} b|^2 - 4\kappa. \tag{4.3}$$

The presence of the free constant -4κ turns out to allow us to find many solutions to (4.3) with a non-trivial $\nabla\!\!\!/ \hat{\otimes} b$. If we can construct solutions with this new type of self-similarity and non-trivial $\nabla\!\!\!/ \hat{\otimes} b$, then the vector field K will still extend to $\{\hat{v} = 0\}$, but now the orbits of K will *not* coincide with null generators of the null hypersurface $\{\hat{v} = 0\}$, and the intrinsic geometry of the $\{\hat{v} = 0\}$ hypersurface will not agree with the intrinsic geometry of a standard null hypersurface in Minkowski space.

Under a suitable smallness assumption on $b|_{v=0}$, one may establish an analogue of the formal expansions of Fefferman–Graham for κSS spacetimes. While a systematic discussion of these formal expansions is outside the scope of the survey article we will observe here some of the consequences of this analysis:

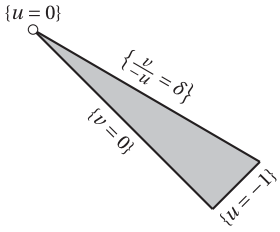


Figure 8. Depiction of first constructed region of exterior of naked singularity.

(1) The best regularity in the $\hat{v}$-variable we can expect for a κSS spacetime is $C^{1,\gamma}$ for $0<\gamma\ll 1$. This is compensated, however, by additional regularity in the u and angular variables as long as we make certain specific (non-trivial) choices of the shift vector b; for any $N\geq 1$, if the shift vector takes a certain specific form along $\{v=0\}$ (to leading order in ϵ) we expect that our spacetime is at least in the class $C^{1,\gamma}_{\hat{v}}C^{N}_{u,\theta^A}$.
(2) A key role is played by developing an existence theory and a priori estimates for equations which are of the form $\psi+\mathscr{L}_X\psi+W\psi=H$, where ψ is an unknown tensor on $\mathbb{S}^2$ and (X,W,H) are given tensors satisfying a suitable smallness condition (cf. (4.3)). It is also important to study the evolutionary analogue of these equations: $(-v)\mathscr{L}_{\partial_v}\psi+\psi+\mathscr{L}_X\psi+W\psi=H$, where now all of the tensors additionally depend on $v\in(0,c)$ for some $c>0$.
(3) As in the choice of $(\mathbf{g}_0,k_0)$ in the Fefferman–Graham expansions, there is a certain set of "free data" which parametrizes the expansions for κSS solutions.

Finally, we note that one may in fact consider versions of the above modification of FGSS where we allow for the form of the regular coordinate $\hat{v}$ to be much more general than $\hat{v}\doteq v^{1-2\kappa}$. We refer to this class of more general spacetimes as "twisted self-similar" (TSS) spacetimes, and we systematically studied them in the work [67] (some specific instances of TSS spacetimes are also important in [62], see Section 4.4 below).

4.3. *The exterior*

In this section we will discuss construction of the *exterior* region of the naked singularity which was carried out in [61]. This is the region to the future of the past light cone of the singularity and is labeled $\mathscr{M}_{\text{ext}}$ in Figure 6.

Let $0<\epsilon\ll\delta\ll 1$. We start by constructing our desired spacetime in the region $\{(v/-u)\in[0,\delta]\}\cap\{u\in(0,-1]\}$ as depictied in Figure 8. While our spacetime will not be an exact κSS spacetime, it will be approximately κSS for $\kappa\sim\epsilon^2$ as $v\to 0$. As with true κSS spacetimes, we will have a regular coordinate system $(\hat{v},u,\theta^A)$ for $\hat{v}\doteq v^{1-2\kappa}$ where the spacetime extends to $\{\hat{v}=0\}$. Tangentially along $\{\hat{v}=0\}$ the spacetime will be exactly κSS in that

$$(v^\kappa\Omega)|_{\hat{v}=0}(u,\theta^A)=(-u)^\kappa,\quad b^A|_{\hat{v}=0}(u,\theta^B)=u^{-1}\tilde{b}^A(\theta^B),\quad \mathbf{g}_{AB}|_{\hat{v}=0}(u,\theta^C)=u^2\tilde{\mathbf{g}}_{AB}(\theta^C),$$

for suitable $\tilde{b}$ and $\tilde{\mathbf{g}}$ so that, in particular, the fundamental constraint Equation (4.3) holds. We also require, further, that along $\{\hat{v}=0\}$, $|b|\sim\epsilon$, and that b takes a certain precise form to leading order as $\epsilon\to 0$.

Our method for constructing this spacetime is to solve a characteristic initial value problem with data along $\{\hat{v}=0\}$ as described in the paragraph above, and data along $\{u=-1\}$ as described below. Then, the construction of the spacetime essentially boils down to suitable a priori estimates. We do not have space in this survey article to discuss all of these, in the following we will only discus some of the ideas involved in the energy estimates for curvature.

In order to keep the presentation here simpler, in our discussion of the outgoing characteristic initial data along $\{u=-1\}$, we will only discuss the choice of $\Omega^{-1}\hat{\chi}$. Our a priori estimates will rely on a renormalization scheme which in turn relies on our spacetime being self-similar to leading order as $(v/-u)\to 0$ (see the discussion below). In particular, it turns out to be necessary that $\Omega^{-1}\hat{\chi}|_{u=-1}=\vartheta|_{u=-1}+O(v^a)$ for suitable $a\gtrsim 1$ and where ϑ must satisfy

$$v\mathscr{L}_{\partial_v}\vartheta_{AB}+\mathscr{L}_{b|_{v=0}}\vartheta_{AB}-(\tfrac{1}{2}\mathrm{div}b|_{v=0}+2\kappa)\vartheta_{AB}-((\nabla\hat{\otimes}b)|_{v=0})^{C}{}_{(A}\vartheta_{B)C}=H_{AB}, \tag{4.4}$$

where H_{AB} is some tensor independent of v, which is determined in terms of $b|_{v=0}$ and satisfies $|H|\sim\epsilon$. We also have that $\kappa\sim\epsilon^2$ and that the operator

$$\vartheta\mapsto\mathscr{L}_{b|_{v=0}}\vartheta_{AB}-\tfrac{1}{2}\mathrm{div}b|_{v=0}\vartheta_{AB}-((\nabla\hat{\otimes}b)|_{v=0})^{C}{}_{(A}\vartheta_{B)C}, \tag{4.5}$$

is anti-symmetric with respect to $L^2(g)$.

Some of the key phenomenon ultimately associated with the solutions ϑ to (4.4) can already be understood if we examine the much simpler scalar equation:

$$v\mathscr{L}_{\partial_v}f-\epsilon^2 f=\epsilon. \tag{4.6}$$

One solution to (4.6) is

$$f=-\frac{1-v^{\epsilon^2}}{\epsilon}.$$

Even though we will have that $f\sim\epsilon^{-1}$ when $v=0$, we also have the bound $|f|\lesssim\epsilon|\log(v)|$. For ϑ we also obtain similar estimates; however, a significant additional difficulty, which is not present in the scalar Equation (4.6), is that we also need to control angular derivatives of ϑ, and naive commutation of the Equation (4.4) destroys the underlying anti-symmetric structure. Thus a careful commutation scheme must be designed.

We now briefly discuss the commutation scheme. The first key fact is that we consider characteristic data which is almost axisymmetric. Thus, the commutation errors with the axisymmetric vector field ∂_ϕ will come with a small constant. The next realization is that one can commute with the operator (4.5). The vector field b turns out to be linearly independent of ∂_ϕ whenever it does not vanish and it thus turns out these commutators suffice to control all derivatives of the solution except at the points where ∂_ϕ or b vanish. (We note, however, that in view of the fact that $|b|\lesssim\epsilon$ the estimates we obtain for higher derivatives have a very bad dependence on ϵ^{-1}.) At the points where ∂_ϕ or b vanish we simply commute with ∇ and the potentially problematic terms that show up turn out to have a good sign.

A key idea to obtain the necessary a priori estimates is to subtract off the expected leading order self-similar behavior of certain curvature and Ricci components before carrying energy, transport, or elliptic estimates. For example, the leading order self-similar behavior of α is generated by a self-similar extension of the ϑ we discussed above. We obtain a new renormalized unknowns $\tilde{\alpha}$ by subtracting off a self-similar extension of ϑ. In particular, with the renormalized unknowns we may use weights are (mildly) singular as $(v/-u)\to 0$. We note that one may anticipate that the error terms generated by this renormalization procedure are small because we know that we can produce formal power series expansions for κSS solutions which are valid as $(v/-u)\to 0$. It is furthermore crucial for this scheme, that even though ϑ is size ϵ^{-1} when $v=0$, it quickly becomes size ϵ once $\log(v)\sim 1$. One particular consequence of these estimates will be that along the hypersurface $(v/-u)=\delta$ all curvature and Ricci components will be small in that (letting Ψ denote a curvature component and ψ a Ricci coefficient)

$$\left|\Psi|_{\frac{v}{-u}=\delta}\right|\lesssim\epsilon|u|^{-2},\quad \left|\psi|_{\frac{v}{-u}=\delta}\right|\lesssim\epsilon|u|^{-1}.$$

The next region we study is the region where $v/-u$ is bounded away from 0 and ∞. The relevant Penrose diagram is depicted in Figure 9. The region covered by the dashed lines corresponds to the previous section, and the light-gray filled in-region is what we will discuss in

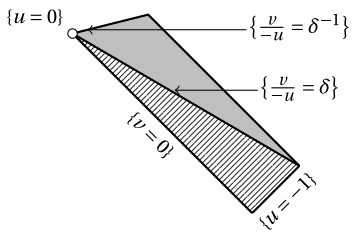


Figure 9. Penrose diagram for second constructed region.

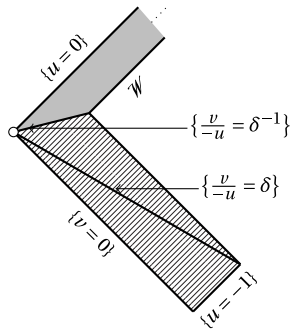


Figure 10. Penrose diagram for the third constructed region.

this section. In this region the estimates are considerably simpler than in the previous region; for example, the key curvature bounds essentially follow from a Grönwell estimate with self-similar weights.

Finally, we come to the construction of the solution up to the hypersurface $\{u = 0\}$. This corresponds to the light gray region in Figure 10 (and the dashed region corresponds to the part we have already constructed).

As a consequences of the of the analysis of the previous region we will already have suitable estimates for all of the double-null quantities along $\{v/-u = \delta^{-1}\}$. We then extend data to an asymptotically flat null hypersurface $\mathscr{W}$. As usual, the key to then constructing the spacetime to the future of $\{v/-u = \delta^{-1}\}$ and $\mathscr{W}$ is to establish suitable a priori estimates. (After constructing the spacetime to the future of $\mathscr{W}$ we can rescale the u-coordinate so that $\mathscr{W}$ corresponds to $\{u = -1\}$.)

This region is, in many ways, dual to the region near $\{v = 0\}$ that we discussed above. We expect the geometry near $\{u = 0\}$ to be modeled on a twisted self-similar solution (see [67]); in particular, we expect to $\Omega\underline{\hat{\chi}}$ to be, at best, Hölder continuous at $\{u = 0\}$ and for $\underline{\alpha}$ to not lie in L^2. In the region near $\{v = 0\}$ there were similar difficulties for $\hat{\chi}$ and α and we solved these via a renormalization scheme which explicitly subtracted off from $\hat{\chi}$ and α a self-similar term generated by the free data ϑ which captured the singular and large parts of $\hat{\chi}$ and α. However, in this region, because we are not posing initial data along $\{u = 0\}$, it is not possible to carry out such an explicit renormalization procedure. Instead, we simply eliminate $\underline{\alpha}$ from our Bianchi system using the renormalized Bianchi equations derived in [49, 50].

4.4. *The interior*

In this section we will discuss construction of the *interior* region of the naked singularity which was carried out in [62]. This is the region to the future of the past light cone of the singularity and

is labeled $\mathscr{M}_{\text{in}t}$ in Figure 6. Our plan will first be to construct the solution in a double-null gauge which holds in a region

$$(u, v, \theta^A) \in \left\{ u \in (-\infty, 0) \text{ and } \frac{v}{-u} \in (-1, 0) \right\}, \tag{4.7}$$

and where the self-similar vector field takes the form $u\partial_u + v\partial_v$. We do not expect this coordinate system to be regular at $\{v = 0\}$ (which will correspond to the past light cone of the singularity) or along the "axis" $\{u = v\}$. In order to carry out this construction we start by identifying a subset of the null-structure equations and use these to derive various equations for the metric components Ω, b, and $\boldsymbol{g}$. (It is outside the scope of this survey article to describe these equations in specific detail.) Using self-similarity we reduce these to a set of equations along the null hypersurface $\{u = -1\}$, and then show the existence of solutions along this hypersurface with suitable boundary conditions at $\{v = -1\}$ and $\{v = 0\}$. After using self-similarity to expand our metric off of $\{u = -1\}$ to the entire region (4.7), the result will be a spacetime where a certain subset of the Einstein vacuum equations hold. We then run a "propagation of constraints" argument to show that the spacetime is actually Ricci flat. Finally, we then construct regular coordinate systems both at $\{v = 0\}$ and at the axis $\{v = u\}$ and show that the interior spacetime may then be glued to our previously constructed exterior spacetime.

Many of the equations we solve will be roughly of the following form[7]:

$$(-v)\partial_v^2\psi + b^A\nabla_A\partial_v\psi + \frac{(-v)A_1}{v+1}\partial_v\psi + A_2\partial_v\psi + \Omega^2(\Delta + A_3(v+1)^{-2})\psi = H, \quad \psi|_{v=0} = h. \tag{4.8}$$

Here $\psi(v, \theta^A) : (-1, 0) \times \mathbb{S}^2 \to \mathbb{R}$ is the unknown, the A_i are suitable constants, H and h can be thought to be given tensors or represent nonlinear terms, $(\Omega, b, \boldsymbol{g})$ will be assumed to satisfy various bootstrap estimates, and Δ is the Laplacian associated to $\boldsymbol{g}$.

A key difficulty is that because the shift vector b does not vanish identically when $v = 0$, our equation is of a mixed elliptic–hyperbolic type. A relatively simple model equation for the elliptic–hyperbolic aspects of the problem (4.8) is

$$(-v)\partial_v^2\psi + \epsilon\partial_\phi\partial_v\psi + \partial_\phi^2\psi = 0, \quad \psi|_{v=-1} = 0, \quad \psi|_{v=0} = h,$$

where our unknown is $\psi(v, \phi) : (-1, 0) \times (-\pi, \pi) \to \mathbb{R}$, we have $0 < \epsilon \ll 1$, and $h : (-\pi, \pi) \to \mathbb{R}$ is given. We consider a suitable family of smoothing operators $\{\Pi_\delta\}_{\delta>0}$ on $\mathbb{S}^2$ which converge to the identity as $\delta \to 0$, and then, for each $0 < \delta \ll 1$, introduce the regularized equation

$$(-v)\partial_v^2\psi_\delta + \epsilon\Pi_\delta(\partial_\phi\partial_v\psi_\delta) + \partial_\phi^2\psi_\delta = 0, \quad \psi_\delta|_{v=-1} = 0, \quad \psi_\delta|_{v=-\delta} = h,$$

for an unknown $\psi_\delta : (-1, -\delta) \times (-\pi, \pi) \to \mathbb{R}$. This equation is elliptic[8] using $-(\epsilon\partial_\phi\psi_\delta + \psi_\delta)$ as a multiplier, one may establish suitable a priori estimates which hold uniformly as $\delta \to 0$. Then it is relatively straightforward to establish the existence of a solution ψ_δ and to obtain the solution ψ as an appropriate limit $\psi_\delta \to_{\delta\to 0} \psi$. Though we will not discuss it in more detail, we note in passing that another important difficulty in the analysis of equations of the form (4.8) is that they are singular near $v = -1$, which corresponds to the breakdown of the double-null foliation at the axis.

In order to construct a good coordinate system near the axis $\{u = v\}$ we introduce self-similar wave coordinates $(\xi_t, \xi_x, \xi_y, \xi_z)$, in the region $\{(v/-u) \in (-1, -\delta)\}$ for suitable $\delta > 0$, which satisfy

$$\Box_g\xi_\cdot = 0, \quad K\xi_\cdot = \xi_\cdot, \tag{4.9}$$

[7] These are relevant both for the original existence argument and for the "propagation of constraints" argument. Also, though we suppress this feature for our survey article, many of the equations considered additionally involve various projections on suitable spaces of spherical harmonics.

[8] Note that the principle symbol of the regularized equation is $(-v)\partial_v^2 + \partial_\phi^2$ and that $-v$ is always positive in $v \in (-1, -\delta)$.

as well as suitable boundary conditions. If we work sufficiently close to the axis, then K will be timelike in the region we are defining $(\xi_t, \xi_x, \xi_y, \xi_z)$ and the system (4.9) reduces to an elliptic equation. Since, as is well-known, the Einstein vacuum equations become quasilinear wave equations in harmonic coordinates, we are eventually able to exploit various elliptic estimates to show that g is smooth in these harmonic coordinates.

Near $\{v = 0\}$ we cannot expect our self-similar wave coordinates to provide a good coordinate system. The solution we construct in the interior region will not directly correspond to the κSS solutions discussed in Section 4.2.2; instead it will be of the more general TSS solutions discussed in [67]. The key good coordinate system turns out to be "lapse renormalized coordinates" $(\hat{v}, u, \theta^A)$ where we define $\hat{v}$ by

$$\hat{v}(v, u, \theta^A) \doteq \int_0^v \Omega^2(\tilde{v}, u, \theta^A)\, \mathrm{d}\tilde{v}.$$

One shows that these coordinates allows for the exterior metric and interior metric to be glued together along $\{\hat{v} = 0\}$ to produce a $C^{1,\gamma}$ metric (with additional regularity in the θ^A and u direction). The resulting spacetime is the desired naked singularity solution.

4.5. *A few open questions*

We close the section by quickly listing some open questions and further directions about naked singularities for the Einstein vacuum questions:

(1) It will be interesting to study both the stability and instability problems of the solutions produced by Theorem 9. Do there exists similar phenomenon to those discovered in Singh's work [19]?
(2) Can one construct naked singularities arising from smooth, or even C^2 initial data?
(3) In higher dimensions, naked singularities are expected to be much more ubiquitous [68, 69]. It would be very interesting to establish existence of asymptotically flat higher dimensional naked singularities and to understand to what extent their presence is generic.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Dedication

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

The author acknowledges support from an Alfred P. Sloan Fellowship in Mathematics and from NSERC discovery grants RGPIN-2021-02562 and DGECR-2021-00093.

Appendix A. Equations of the double-null gauge

In this appendix we collect the equations of the double-null gauge for ease of reference and to fix the notation. For a proper introduction to the double-null gauge we refer the reader to the works [10, 33]. We will say that $(\mathscr{M}, g)$ is in the double-null gauge with the shift in the u-direction if

$$g = -2\Omega^2(\mathrm{d}u \otimes \mathrm{d}v + \mathrm{d}v \otimes \mathrm{d}u) + \not{g}_{AB}(\mathrm{d}\theta^A - b^A\,\mathrm{d}u) \otimes (\mathrm{d}\theta^B - b^B\,\mathrm{d}u), \tag{A.1}$$

holds, and that $(\mathscr{M}, g)$ is in the double-null gauge with the shift in the v-direction if (A.1) holds with the roles of u and v interchanged. In the following, unless indicated otherwise, we assume that the shift in the u-direction.

We define the null vector fields e_4 and e_3 by

$$e_4 \doteq \Omega^{-1}\partial_v, \quad e_3 \doteq \Omega^{-1}(\partial_u + b).$$

The $\mathbb{S}^2_{u,v}$-projected derivatives in the e_4 and e_3 direction are denoted by ∇_4 and ∇_3. The induced covariant derivative along each $\mathbb{S}^2_{u,v}$ is denoted by $\not{\nabla}$, $*$ refers to the Hodge star operator on $(\mathbb{S}^2, \not{g})$, slashed operators are defined with $\not{g}$, and we use K to denote the Gaussian curvature of $\not{g}$. We will use $\mathscr{L}$ to denote the Lie derivative.

The Ricci coefficients $(\omega, \underline{\omega}, \zeta_A, \eta_A, \underline{\eta}_A, \chi_{AB}, \underline{\chi}_{AB})$ satisfy

$$\begin{gathered}
\omega = -\tfrac{1}{2}\nabla_4 \log\Omega, \quad \underline{\omega} = -\tfrac{1}{2}\nabla_3 \log\Omega, \quad \mathscr{L}_{\partial_v} b^A = -4\Omega^2\zeta^A, \\
\eta_A = \zeta_A + \not{\nabla}_A \log\Omega, \quad \underline{\eta}_A = -\zeta_A + \not{\nabla}_A \log\Omega, \\
\mathscr{L}_{e_4}\not{g}_{AB} = \chi_{AB}, \quad \mathscr{L}_{e_3}\not{g}_{AB} = \underline{\chi}_{AB}.
\end{gathered} \tag{A.2}$$

We split χ_{AB} and $\underline{\chi}_{AB}$ into their trace and trace-free parts via

$$\chi_{AB} = \hat{\chi}_{AB} + \tfrac{1}{2}\mathrm{tr}\chi\not{g}_{AB}, \quad \underline{\chi}_{AB} = \hat{\underline{\chi}}_{AB} + \tfrac{1}{2}\mathrm{tr}\underline{\chi}\not{g}_{AB}.$$

The null curvature components $(\alpha_{AB}, \beta_A, \rho, \sigma, \underline{\beta}_A, \underline{\alpha}_{AB})$ a re related to the Ricci coefficients via the various null-structure equations:

$$\nabla_4 \mathrm{tr}\chi + \tfrac{1}{2}(\mathrm{tr}\chi)^2 = -|\hat{\chi}|^2 - 2\omega\mathrm{tr}\chi, \tag{A.3}$$

$$\nabla_4 \hat{\chi} + \mathrm{tr}\chi\hat{\chi} = -\alpha - 2\omega\hat{\chi},$$

$$\nabla_3 \mathrm{tr}\underline{\chi} + \tfrac{1}{2}(\mathrm{tr}\underline{\chi})^2 = -|\hat{\underline{\chi}}|^2 - 2\underline{\omega}\mathrm{tr}\underline{\chi}, \tag{A.4}$$

$$\nabla_3 \hat{\underline{\chi}} + \mathrm{tr}\underline{\chi}\hat{\underline{\chi}} = -\underline{\alpha} - 2\underline{\omega}\hat{\underline{\chi}},$$

$$\nabla_3 \hat{\chi} + \tfrac{1}{2}\mathrm{tr}\underline{\chi}\hat{\chi} = 2\underline{\omega}\hat{\chi} + \not{\nabla}\hat{\otimes}\eta + \eta\hat{\otimes}\eta - \tfrac{1}{2}\mathrm{tr}\chi\hat{\underline{\chi}}, \tag{A.5}$$

$$\nabla_3 \mathrm{tr}\chi + \tfrac{1}{2}\mathrm{tr}\chi\mathrm{tr}\underline{\chi} = 2\rho + 2\underline{\omega}\mathrm{tr}\chi + 2\not{\mathrm{div}}\eta + 2|\eta|^2 - \hat{\chi}\cdot\hat{\underline{\chi}}, \tag{A.6}$$

$$\nabla_4 \hat{\underline{\chi}} + \tfrac{1}{2}\mathrm{tr}\chi\hat{\underline{\chi}} = 2\omega\hat{\underline{\chi}} + \not{\nabla}\hat{\otimes}\underline{\eta} + \underline{\eta}\hat{\otimes}\underline{\eta} - \tfrac{1}{2}\mathrm{tr}\underline{\chi}\hat{\chi},$$

$$\nabla_4 \mathrm{tr}\underline{\chi} + \tfrac{1}{2}\mathrm{tr}\chi\mathrm{tr}\underline{\chi} = 2\rho + 2\omega\mathrm{tr}\underline{\chi} + 2\not{\mathrm{div}}\underline{\eta} + 2|\underline{\eta}|^2 - \hat{\chi}\cdot\hat{\underline{\chi}},$$

$$\nabla_4 \eta = -\chi\cdot(\eta - \underline{\eta}) - \beta,$$

$$\nabla_3 \underline{\eta} = -\underline{\chi}\cdot(\underline{\eta} - \eta) + \underline{\beta},$$

$$\not{\mathrm{curl}}\eta = \sigma + \tfrac{1}{2}\hat{\underline{\chi}}\wedge\hat{\chi},$$

$$\not{\mathrm{curl}}\underline{\eta} = -\sigma - \tfrac{1}{2}\hat{\underline{\chi}}\wedge\hat{\chi},$$

$$\nabla_4 \underline{\omega} = \tfrac{1}{2}\rho + 2\underline{\omega}\omega + \tfrac{1}{2}|\eta|^2 - \eta\cdot\underline{\eta},$$

$$\nabla_3 \omega = \tfrac{1}{2}\rho + 2\underline{\omega}\omega + \tfrac{1}{2}|\underline{\eta}|^2 - \eta\cdot\underline{\eta},$$

$$K = -\rho + \tfrac{1}{2}\hat{\chi}\cdot\hat{\underline{\chi}} - \tfrac{1}{4}\mathrm{tr}\chi\mathrm{tr}\underline{\chi},$$

$$\not{\mathrm{div}}\hat{\chi} - \tfrac{1}{2}\not{\nabla}\mathrm{tr}\chi = -\beta + \tfrac{1}{2}\mathrm{tr}\chi\zeta - \zeta\cdot\hat{\chi},$$

$$\not{\mathrm{div}}\hat{\underline{\chi}} - \tfrac{1}{2}\not{\nabla}\mathrm{tr}\underline{\chi} = \underline{\beta} - \tfrac{1}{2}\mathrm{tr}\underline{\chi}\zeta + \zeta\cdot\hat{\underline{\chi}},$$

Finally we have the Bianchi equations for the curvature components:

$$\nabla_3\alpha + \tfrac{1}{2}\mathrm{tr}\underline{\chi}\alpha = \nabla\hat{\otimes}\beta + 4\underline{\omega}\alpha - 3(\hat{\chi}\rho + {}^*\hat{\chi}\sigma) + (\zeta + 4\eta)\hat{\otimes}\beta,$$

$$\nabla_4\beta + 2\mathrm{tr}\chi\beta = \mathrm{div}\alpha - 2\omega\beta + (2\zeta + \underline{\eta})\cdot\alpha,$$

$$\nabla_3\beta + \mathrm{tr}\underline{\chi}\beta = \nabla\rho + 2\underline{\omega}\beta + {}^*\nabla\sigma + 2\hat{\chi}\cdot\underline{\beta} + 3(\eta\rho + {}^*\eta\sigma),$$

$$\nabla_4\sigma + \tfrac{3}{2}\mathrm{tr}\chi\sigma = -\mathrm{div}^*\beta + \tfrac{1}{2}\hat{\underline{\chi}}\wedge\alpha - \zeta\wedge\beta - 2\underline{\eta}\wedge\beta,$$

$$\nabla_4\rho + \tfrac{3}{2}\mathrm{tr}\chi\rho = \mathrm{div}\beta - \tfrac{1}{2}\hat{\underline{\chi}}\cdot\alpha + \zeta\cdot\beta - 2\eta\cdot\beta$$

$$\nabla_3\sigma + \tfrac{3}{2}\mathrm{tr}\underline{\chi}\sigma = -\mathrm{div}^*\underline{\beta} - \tfrac{1}{2}\hat{\chi}\wedge\underline{\alpha} + \zeta\wedge\underline{\beta} - 2\eta\wedge\underline{\beta},$$

$$\nabla_3\rho + \tfrac{3}{2}\mathrm{tr}\underline{\chi}\rho = -\mathrm{div}\underline{\beta} - \tfrac{1}{2}\hat{\chi}\cdot\underline{\alpha} + \zeta\cdot\underline{\beta} + 2\underline{\eta}\cdot\underline{\beta},$$

$$\nabla_4\underline{\beta} + \mathrm{tr}\chi\underline{\beta} = -\nabla\rho + 2\omega\underline{\beta} + {}^*\nabla\sigma + 2\hat{\underline{\chi}}\cdot\beta - 3(\underline{\eta}\rho - {}^*\underline{\eta}\sigma),$$

$$\nabla_3\underline{\beta} + 2\mathrm{tr}\underline{\chi}\underline{\beta} = -\mathrm{div}\underline{\alpha} - 2\underline{\omega}\underline{\beta} - (-2\zeta + \eta)\cdot\underline{\alpha},$$

$$\nabla_4\underline{\alpha} + \tfrac{1}{2}\mathrm{tr}\chi\underline{\alpha} = -\nabla\hat{\otimes}\underline{\beta} + 4\omega\underline{\alpha} - 3(\hat{\underline{\chi}}\rho - {}^*\hat{\underline{\chi}}\sigma) - (-\zeta + 4\underline{\eta})\hat{\otimes}\underline{\beta}.$$

Appendix B. The renormalized Bianchi equations from [49, 50]

We list here the renormalized Bianchi Equations from [49, 50]. We first define

$$\check{\rho} \doteq \rho - \tfrac{1}{2}\hat{\chi}\cdot\hat{\underline{\chi}}, \quad \check{\sigma} \doteq \sigma - \tfrac{1}{2}\hat{\chi}\wedge\hat{\underline{\chi}}.$$

Then we have

$$\begin{aligned}\nabla_3\beta + \mathrm{tr}\underline{\chi}\beta =\ & \nabla\check{\rho} + {}^*\nabla\check{\sigma} + 2\underline{\omega}\beta + 2\hat{\chi}\cdot\underline{\beta} + 3(\eta\check{\rho} + {}^*\eta\check{\sigma})\\ & + \tfrac{1}{2}(\nabla(\hat{\chi}\cdot\hat{\underline{\chi}}) + {}^*\nabla(\hat{\chi}\wedge\hat{\underline{\chi}})) + \tfrac{3}{2}(\eta\hat{\chi}\cdot\hat{\underline{\chi}} + {}^*\eta\hat{\chi}\wedge\hat{\underline{\chi}}),\end{aligned}$$

$$\nabla_4\check{\sigma} + \tfrac{3}{2}\mathrm{tr}\chi\check{\sigma} = -\mathrm{div}^*\beta - \zeta\wedge\beta - 2\underline{\eta}\wedge\beta + \tfrac{1}{2}\hat{\chi}\wedge\left(\nabla\hat{\otimes}\underline{\eta}\right) - \tfrac{1}{2}\hat{\chi}\wedge(\underline{\eta}\hat{\otimes}\underline{\eta}),$$

$$\nabla_4\check{\rho} + \tfrac{3}{2}\mathrm{tr}\chi\check{\rho} = \mathrm{div}\beta + \zeta\cdot\beta + 2\underline{\eta}\cdot\beta - \tfrac{1}{2}\hat{\chi}\cdot(\nabla\hat{\otimes}\underline{\eta}) - \tfrac{1}{2}\hat{\chi}\cdot(\underline{\eta}\hat{\otimes}\underline{\eta}) + \tfrac{1}{4}\mathrm{tr}\underline{\chi}|\hat{\chi}|^2.$$

$$\nabla_3\check{\sigma} + \tfrac{3}{2}\mathrm{tr}\underline{\chi}\check{\sigma} = \mathrm{div}^*\underline{\beta} + \zeta\wedge\underline{\beta} - 2\eta\wedge\underline{\beta} + \tfrac{1}{2}\hat{\underline{\chi}}\wedge(\nabla\hat{\otimes}\eta) + \tfrac{1}{2}\hat{\underline{\chi}}(\eta\hat{\otimes}\eta),$$

$$\nabla_3\check{\rho} + \tfrac{3}{2}\mathrm{tr}\underline{\chi}\check{\rho} = -\mathrm{div}\underline{\beta} + \zeta\cdot\underline{\beta} - 2\eta\cdot\underline{\beta} - \tfrac{1}{2}\hat{\underline{\chi}}\cdot(\nabla\hat{\otimes}\eta) - \tfrac{1}{2}\hat{\underline{\chi}}\cdot(\eta\hat{\otimes}\eta) + \tfrac{1}{2}\mathrm{tr}\chi|\hat{\underline{\chi}}|^2,$$

$$\begin{aligned}\nabla_4\underline{\beta} + \mathrm{tr}\chi\underline{\beta} =\ & -\nabla\check{\rho} + {}^*\nabla\check{\sigma} + 2\omega\underline{\beta} + 2\hat{\underline{\chi}}\cdot\beta - 3(\underline{\eta}\check{\rho} - {}^*\underline{\eta}\check{\sigma})\\ & - \tfrac{1}{2}(\nabla(\hat{\underline{\chi}}\cdot\hat{\chi}) + {}^*\nabla(\hat{\underline{\chi}}\wedge\hat{\chi})) - \tfrac{3}{2}(\underline{\eta}\hat{\underline{\chi}}\wedge\hat{\chi} + {}^*\underline{\eta}\hat{\underline{\chi}}\wedge\hat{\chi}).\end{aligned}$$

Appendix C. Characteristic initial value problem

In this section we will quickly review the well-posedness of the characteristic initial value problem for the Einstein equations. Choose real numbers u_0, u_1, v_0, and v_1 with $u_0 < u_1$ and $v_0 < v_1$. We say a triple of $\left(\Omega^{(\mathrm{in})}, \left(b^A\right)^{(\mathrm{in})}, \boldsymbol{g}^{(\mathrm{in})}_{AB}\right)$ defined along $[u_0, u_1]\times\mathbb{S}^2$, a pair $\left(\Omega^{(\mathrm{out})}, \boldsymbol{g}^{(\mathrm{out})}_{AB}\right)$ defined along $[v_0, v_1]\times\mathbb{S}^2$, and a 1-form $\tilde{\zeta}_A$ on $\mathbb{S}^2$ form a characteristic initial data set if

(1) We have $\left(\Omega^{(\mathrm{in})}, \boldsymbol{g}^{(\mathrm{in})}\right)|_{u=u_0} = \left(\Omega^{(\mathrm{out})}, \boldsymbol{g}^{(\mathrm{out})}\right)|_{v=v_0}$.

(2) We consider $\left(\Omega^{(\mathrm{in})}, \left(b^A\right)^{(\mathrm{in})}, \boldsymbol{g}^{(\mathrm{in})}_{AB}\right)$ as being defined along a hypersurface $\{v = v_0\}$ and $\left(\Omega^{(\mathrm{out})}, \boldsymbol{g}^{(\mathrm{out})}_{AB}\right)$ as being defined along a hypersurface $\{u = u_0\}$. After defining $\underline{\omega}$, ω, χ, and $\underline{\chi}$ using (A.2), we require that the Raychaudhuri Equations (A.3) and (A.6) hold along $\{u = u_0\}$ and $\{v = v_0\}$ respectively.

We can now state a local existence theorem.

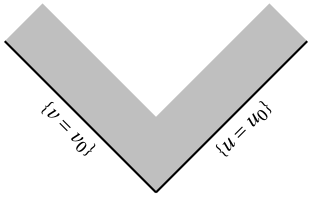


Figure 11. Penrose diagram for spacetime arising from characteristic initial value problem.

Theorem 12. *[70, 71] Let* $\left(\Omega^{(\mathrm{in})}, \left(b^A\right)^{(\mathrm{in})}, \boldsymbol{g}_{AB}^{(\mathrm{in})}\right)$, $\left(\Omega^{(\mathrm{out})}, \boldsymbol{g}_{AB}^{(\mathrm{out})}\right)$, *and* $\tilde{\zeta}_A$ *form an initial data set of sufficiently high regularity. Then there exists a constant* $c > 0$ *and a unique solution* $\left(\mathcal{M}, g\right)$ *to the Einstein vacuum equations expressed in a double-null gauge and defined in a region* $\left(u, v, \theta^A\right) \in \left([u_0, u_1] \times [v_0, v_0 + c] \times \mathbb{S}^2\right) \cup \left([u_0, u_0 + c] \times [v_0, v_1] \times \mathbb{S}^2\right)$ *so that*

$$\left(\Omega^{(\mathrm{in})}, \left(b^A\right)^{(\mathrm{in})}, \boldsymbol{g}_{AB}^{(\mathrm{in})}\right) = \left(\Omega, b^A, \boldsymbol{g}_{AB}\right)|_{v=v_0}, \quad \left(\Omega^{(\mathrm{out})}, \boldsymbol{g}_{AB}^{(\mathrm{out})}\right) = \left(\Omega, \boldsymbol{g}_{AB}\right)|_{u=u_0}$$
$$\tilde{\zeta}_A = \zeta_A|_{(u,v)=(u_0,v_0)}.$$

In Figure 11 we have drawn a Penrose diagram depicting the range of the solution.

References

[1] Y. Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), pp. 141–225.

[2] Y. Choquet-Bruhat and R. P. Geroch, "Global aspects of the Cauchy problem in general relativity", *Commun. Math. Phys.* **14** (1969), pp. 329–335.

[3] R. M. Wald, *General Relativity*, The University of Chicago Press: Chicago, 1984.

[4] K. Schwarzschild, "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie", *Sitzungsber. K. Preuss. Akad. Wiss.* **1** (1916), pp. 189–196.

[5] M. Dafermos and I. Rodnianski, "Lectures on black holes and linear waves", in *Evolution Equations*, Clay Mathematics Proceedings, American Mathematical Society: Providence, RI, 2013, pp. 97–205.

[6] J. Sbierski, "The C_0-inextendibility of the Schwarzschild spacetime and the spacelike diameter in Lorentzian geometry", *J. Differ. Geom.* **108** (2018), no. 2, pp. 319–378.

[7] J. Sbierski, "On the proof of the C^0-inextendibility of the Schwarzschild spacetime", *J. Phys. Conf. Ser.* **968** (2018), article no. 012012.

[8] R. Penrose, "Gravitational collapse and space-time singularities", *Phys. Rev. Lett.* **14** (1965), pp. 57–59.

[9] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series, Princeton University Press: Princeton, NJ, 1993, pp. x+514.

[10] S. Klainerman and F. Nicolò, *The Evolution Problem in General Relativity*, Progress in Mathematical Physics, Birkhäuser Boston Inc.: Boston, MA, 2003, pp. xiv+385.

[11] D. Christodoulou, "On the global initial value problem and the issue of singularities", *Class. Quantum Gravity* **16** (1999), A23–A35.

[12] D. Shen, *Global stability of Minkowski spacetime with minimal decay*, preprint, 2023, 2310.07483.

[13] R. Penrose, "Gravitational collapse: the role of general relativity", *Riv. Nuovo Cim.* **1** (1969), pp. 252–276.

[14] D. Christodoulou, "Examples of naked singularity formation in the gravitational collapse of a scalar field", *Ann. Math. (2)* **140** (1994), no. 3, pp. 607–653.

[15] M. Dafermos and J. Luk, "The interior of dynamical vacuum black holes I: the C^0-stability of the Kerr Cauchy horizon", *Ann. Math.* (2017). (to appear).

[16] D. Christodoulou, "The formation of black holes and singularities in spherically symmetric gravitational collapse", *Commun. Pure Appl. Math.* **44** (1991), no. 3, pp. 339–373.

[17] D. Christodoulou, "Bounded variation solutions of the spherically symmetric Einstein-scalar field equations", *Commun. Pure Appl. Math.* **46** (1993), no. 8, pp. 1131–1220.

[18] D. Christodoulou, "The instability of naked singularities in the gravitational collapse of a scalar field", *Ann. Math.* **149** (1999), pp. 183–217.

[19] J. Singh, *High regularity waves on self-similar naked singularity interiors: decay and the role of blue-shift*, preprint, 2024, 2402.00062.

[20] M. Dafermos, “Spherically symmetric spacetimes with a trapped surface”, *Class. Quantum Gravity* **22** (2005), no. 11, pp. 2221–2232.

[21] J. Singh, “A construction of approximately self-similar naked singularities for the spherically symmetric Einstein-scalar field system”, *Ann. Henri Poincaré* (2024).

[22] D. Christodoulou, “Violation of cosmic censorship in the gravitational collapse of a dust cloud”, *Commun. Math. Phys.* **93** (1984), no. 2, pp. 171–195.

[23] Y. Guo, M. Hadzic and J. Jang, “Naked singularities in the Einstein–Euler system”, *Ann. PDE* **9** (2023), no. 1, article no. 4.

[24] P. Bizoń and A. Wasserman, “On the existence of self-similar spherically symmetric wave maps coupled to gravity”, *Class. Quantum Gravity* **19** (2002), no. 12, pp. 3309–3321.

[25] M. Choptuik, “Universality and scaling in gravitational collapse of a massive scalar field”, *Phys. Rev. Lett.* **70** (1993), pp. 9–12.

[26] C. Gundlach and J. Martin-Garcia, “Critical phenomena in gravitational collapse”, *Living Rev. Relativ.* **10** (2007), article no. 5.

[27] J. Liu and J. Li, “A robust proof of the instability of naked singularities of a scalar field in spherical symmetry”, *Commun. Math. Phys.* **363** (2018), no. 2, pp. 561–578.

[28] J. Li and J. Liu, “Instability of spherical naked singularities of a scalar field under gravitational perturbations”, *J. Differ. Geom.* **120** (2022), no. 1, pp. 97–197.

[29] X. An, “Naked singularity censoring with aniostropic apparent horizon”, *Ann. Math.* (2024). (to appear).

[30] V. Cardoso, J. A. L. Costa, K. Destounis, P. Hintz and A. Jansen, “Quasinormal modes and strong cosmic censorship”, *Phys. Rev. Lett.* **120** (2018), article no. 031103.

[31] M. Dafermos and Y. Shlapentokh-Rothman, “Rough initial data and the strength of the blue-shift instability on cosmological black holes with Λ ensuremath> 0”, *Class. Quantum Gravity* **35** (2018), no. 19, article no. 195010.

[32] O. J. C. Dias, H. S. Reall and J. E. Santos, “Strong cosmic censorship: taking the rough with the smooth”, *J. High Energy Phys.* (2018), no. 10, article no. 1.

[33] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society (EMS): Zürich, 2009, pp. x+589.

[34] X. Chen and S. Klainerman, *Formation of trapped surfaces in geodesic foliation*, preprint, 2024, 2409.14582.

[35] M. Dafermos, “The formation of black holes in general relativity [after D. Christodoulou]”, *Sémin. Bourbaki* **2011/2012** (2013), no. 352, pp. 243–313. Exposés 1043–1058, p. Exp. No. 1051, viii.

[36] J. Li and P. Yu, “Construction of Cauchy data of vacuum Einstein field equations evolving to black holes”, *Ann. Math. (2)* **181** (2015), no. 2, pp. 699–768.

[37] J. Corvino and R. M. Schoen, “On the asymptotics for the vacuum Einstein constraint equations”, *J. Differ. Geom.* **73** (2006), no. 2, pp. 185–217.

[38] C. Kehle and R. Unger, “Event horizon gluing and black hole formation in vacuum: the very slowly rotating case”, *Adv. Math.* **452** (2024), article no. 109816.

[39] S. Klainerman and I. Rodnianski, “On the formation of trapped surfaces”, *Acta Math.* **208** (2012), no. 2, pp. 211–333.

[40] S. Klainerman and I. Rodnianski, “On emerging scarred surfaces for the Einstein vacuum equations”, *Discrete Contin. Dyn. Syst.* **28** (2010), no. 3, pp. 1007–1031.

[41] X. An, *Formation of trapped surfaces in general relativity*, PhD thesis, Princeton University: ProQuest LLC, Ann Arbor, MI, 2014.

[42] S. Klainerman, J. Luk and I. Rodnianski, “A fully anisotropic mechanism for formation of trapped surfaces in vacuum”, *Invent. Math.* **198** (2014), no. 1, pp. 1–26.

[43] S. Klainerman, I. Rodnianski and J. Szeftel, “The bounded L^2 curvature conjecture”, *Invent. Math.* **202** (2015), no. 1, pp. 91–216.

[44] J. Szeftel, “Sharp Strichartz estimates for the wave equation on a rough background”, *Ann. Sci. Éc. Norm. Supér. (4)* **49** (2016), no. 6, pp. 1279–1309.

[45] J. Szeftel, “Parametrix for wave equations on a rough background III: space–time regularity of the phase”, *Astérisque* (2018), no. 401, pp. viii+321.

[46] J. Szeftel, “Parametrix for wave equations on a rough background. I: regularity of the phase at initial time. II: construction and control at initial time”, *Astérisque* (2023), no. 443, pp. ix+275.

[47] J. Szeftel, “Parametrix for wave equations on a rough background: IV. Control of the error term”, *Astérisque* (2023), no. 444, pp. viii+314.

[48] J. Luk, “Singularities in general relativity”, in *ICM—International Congress of Mathematicians. Vol. 5. Sections 9–11*, EMS Press: Berlin, 2023, pp. 4120–4141. ©2023.

[49] J. Luk and I. Rodnianski, “Local propagation of impulsive gravitational waves”, *Commun. Pure Appl. Math.* **68** (2015), no. 4, pp. 511–624.

[50] J. Luk and I. Rodnianski, "Nonlinear interaction of impulsive gravitational waves for the vacuum Einstein equations", *Camb. J. Math.* **5** (2017), no. 4, pp. 435–570.
[51] R. Penrose, "The geometry of impulsive gravitational waves", in *General Relativity (Papers in Honour of J. L. Synge)*, Clarendon Press: Oxford, 1972, pp. 101–115.
[52] K. A. Khan and R. Penrose, "Scattering of two impulsive gravitational plane waves", *Nature* **229** (1971), no. 5281, pp. 185–186.
[53] J. Luk and I. Rodnianski, *High-frequency limits and null dust solutions in general relativity*, preprint, 2020, 2009.08968.
[54] G. A. Burnett, "The high-frequency limit in general relativity", *J. Math. Phys.* **30** (1989), no. 1, pp. 90–96.
[55] C. Huneau and J. Luk, "High-frequency solutions to the Einstein equations", *Class. Quantum Gravity* **41** (2024), article no. 143002.
[56] X. An and J. Luk, "Trapped surfaces in vacuum arising dynamically from mild incoming radiation", *Adv. Theor. Math. Phys.* **21** (2017), no. 1, pp. 1–120.
[57] X. An, "Emergence of apparent horizon in gravitational collapse", *Ann. PDE* **6** (2020), no. 2, article no. 10.
[58] X. An, "A scale-critical trapped surface formation criterion: a new proof via signature for decay rates", *Ann. PDE* **8** (2022), no. 1, article no. 3.
[59] X. An and T. He, "Dynamics of apparent horizon and a null comparison principle", *Ann. PDE* **10** (2024), article no. 15.
[60] X. An and Q. Han, *Anisotropic dynamical horizons arising in gravitational collapse*, preprint, 2020, 2010.12524.
[61] I. Rodnianski and Y. Shlapentokh-Rothman, "Naked singularities for the Einstein vacuum equations: the exterior solution", *Ann. Math. (2)* **198** (2023), no. 1, pp. 231–391.
[62] Y. Shlapentokh-Rothman, *Naked singularities for the Einstein vacuum equations: the interior solution*, preprint, 2022, 2204.09891.
[63] F. J. Tipler, "Singularities in conformally flat spacetimes", *Phys. Lett. A* **64** (1977), no. 1, pp. 8–10.
[64] C. Fefferman and C. R. Graham, "Conformal invariants", *Astérisque* (1985), pp. 95–116. Numéro Hors Série The Mathematical Heritage of Élie Cartan (Lyon, 1984).
[65] C. Fefferman and C. R. Graham, *The Ambient Metric*, Annals of Mathematics Studies, Princeton University Press: Princeton, NJ, 2012, pp. x+113.
[66] I. Rodnianski and Y. Shlapentokh-Rothman, "The asymptotically self-similar regime for the Einstein vacuum equations", *Geom. Funct. Anal.* **28** (2018), no. 3, pp. 755–878.
[67] Y. Shlapentokh-Rothman, "Twisted self-similarity and the Einstein vacuum equations", *Commun. Math. Phys.* **401** (2023), no. 2, pp. 2269–2325.
[68] L. Lehner and F. Pretorius, "Black strings, low viscosity fluids, and violation of cosmic censorship", *Phys. Rev. Lett.* **105** (2010), no. 10, article no. 101102.
[69] X. An and X. Zhang, "Examples of naked singularity formation in higher-dimensional Einstein-vacuum spacetimes", *Ann. Henri Poinc.* **19** (2018), no. 2, pp. 619–651.
[70] A. D. Rendall, "Reduction of the characteristic initial value problem to the Cauchy problem and its applications to the Einstein equations", *Proc. R. Soc. Lond. A* **427** (1990), no. 1872, pp. 221–239.
[71] J. Luk, "On the local existence for the characteristic initial value problem in general relativity", *Int. Math. Res. Not. IMRN* (2012), no. 20, pp. 4625–4678.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 415-454
https://doi.org/10.5802/crmeca.271

Review article / *Article de synthèse*

The Strong Cosmic Censorship conjecture

La conjecture de Censure Cosmique forte

Maxime Van de Moortel[a]

[a] Department of Mathematics, Rutgers University, Hill Center, New Brunswick NJ 08854, USA
E-mail: maxime.vandemoortel@rutgers.edu

Abstract. In the wake of major breakthroughs in General Relativity during the 1960s, Roger Penrose introduced Strong Cosmic Censorship, a profound conjecture regarding the deterministic nature of the theory. Penrose's proposal has since opened far-reaching new mathematical avenues, revealing connections to fundamental questions about black holes and the nature of gravitational singularities. We review recent advances arising from modern techniques in the theory of partial differential equations as applied to Strong Cosmic Censorship, maintaining a focus on the context of gravitational collapse that gave birth to the conjecture.

Résumé. Dans le sillage des percées majeures de la relativité générale dans les années 1960, Roger Penrose a introduit la censure cosmique forte, une conjecture profonde concernant la nature déterministe de la théorie. La proposition de Penrose a depuis ouvert de nouvelles voies mathématiques d'une grande portée, révélant des liens avec des questions fondamentales sur les trous noirs et la nature des singularités gravitationnelles. Nous passons en revue les avancées récentes issues des techniques modernes de la théorie des équations aux dérivées partielles appliquées à la censure cosmique forte, en nous concentrant sur le contexte de l'effondrement gravitationnel qui a donné naissance à la conjecture.

Keywords. General relativity, Laplacian determinism, Einstein equations, Black holes.

Mots-clés. Relativité générale, Déterminisme de Laplace, Équations d'Einstein, Trous noirs.

Funding. NSF Grant DMS-2247376.

Manuscript received 9 October 2024, accepted 10 October 2024.

> The proving (or disproving) of [...] Cosmic Censorship in classical general relativity is perhaps the most important unsolved problem in that theory.
>
> Penrose (1982), [1]

1. Introduction

1.1. *Cosmic Censorship: from the absence of naked singularities to determinism*

Penrose held the Cosmic Censorship hypothesis in such high regard because, in his words, "unlike many other problems", it has "genuine astrophysical significance" [1]. Cosmic Censorship is indeed inextricably tied to the very question of Laplacian determinism in the general theory of relativity; namely, is "the present state of the universe the effect of its past and the cause of its future" [2]? Originally, the idea of Cosmic Censorship broadly asserts that should singularities exist within general relativity, they ought to be confined inside a black hole and, thus, do not

affect the observable universe[1]. In other words, a cosmic censor conceals any *naked* (that is to say, visible to infinitely far-away observers) singularity within a black hole [4]. Very early on, however, Penrose motivated cosmic censorship as a way to avoid "the unpredictability entailed by the presence of naked singularities" [5]. In modern terms, the occurrence of a naked singularity is indeed responsible for the formation of a Cauchy horizon (see [6,7] for a precise definition) emanating from it. Should this Cauchy horizon be a regular hypersurface, the trajectory of observers crossing it is no longer determined by their initial state. In Penrose's words: "there being no theory governing what happens as the result of the appearance of such a singularity, this particular observer would not be able to account, in scientific terms, for whatever physical behaviour is seen" [8]. In this situation, the future of spacetime itself beyond the Cauchy horizon cannot, in fact, be predicted from its initial state, in patent violation of the spirit of Laplacian determinism; such considerations will be discussed with greater depth in Section 2.

1.2. *Locally naked singularities and early formulations of Strong Cosmic Censorship*

Penrose realized that a breakdown of Laplacian determinism would also occur if the naked singularity was covered by a black hole, an instance he calls "a locally naked singularity", namely a singularity that is still visible to observers falling into the black hole [5]. In his words, "cosmic censorship [...] should [..] preclude such locally naked singularities" [5], a realization which prompted him to introduce two distinct notions of cosmic censorship [9]:

- the absence of naked singularities in the process of gravitational collapse (the original cosmic censorship [4]), dubbed *Weak Cosmic Censorship* by subsequent authors [10,11].
- the absence of naked or locally naked singularities (possibly inside a black hole) in generic gravitational collapse, called by Penrose "*Strong Cosmic Censorship*" [5,9].

The terminology is fully justified in these original forms because the Strong Cosmic Censorship statement implies that of Weak Cosmic Censorship (see [12], section 5 for a historical perspective on this, but already contrast with the modern formulation of Strong Cosmic Censorship in our Section 1.4 below). While originally, Cosmic Censorship concerned the problem of naked singularities globally affecting the universe (especially idealized observers located at infinity), the change of paradigm in Penrose's work [5,9] refocused Cosmic Censorship on a failure of Laplacian determinism, which can occur locally, with no reference to the observable universe in its globality.

1.3. *The interior of the Kerr black hole and the issue of Cauchy horizons*

The interior of the (sub-extremal) Kerr metric (representing, at the time of Penrose, the most appropriate known model for an astrophysical black hole), which also features a Cauchy horizon, provides a striking case study. Contrary to Cauchy horizons emanating from a naked singularity, the Kerr *Cauchy horizon exists*[2] *despite the absence of any singularity* in its past. In retrospect, this fact demonstrates that the study of determinism (or its breakdown) is not directly tied to naked singularities and falls under more general considerations. That said, the Kerr metric interior (analytically extended to the future of its Cauchy horizon) also features a timelike singularity, which is locally naked from the viewpoint of in-falling observers crossing the Cauchy horizon, see Figure 1. Penrose found such timelike singularities to be the most physically problematic issue

[1] Before the 1960's, relativistic singularities were viewed as absurdities that could, however, be ignored providing they were located inside a black hole, which was not then considered as a physically realistic spacetime, see [3], Chapter 3.2.

[2] The Kerr Cauchy horizon coincides with its inner horizon, and emanates from timelike infinity. This is another common mechanism for the appearance of Cauchy horizons, see [13,14].

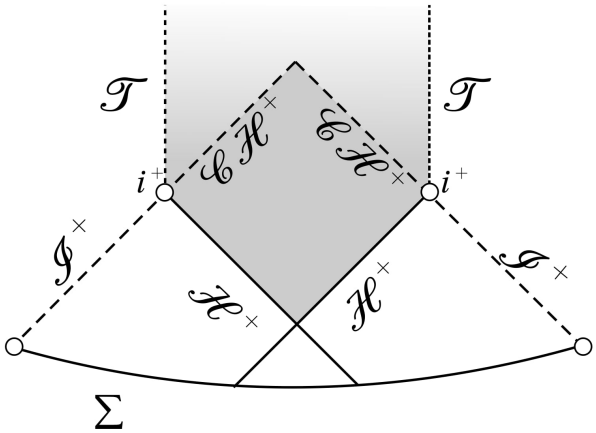


Figure 1. Future analytic extension of the Kerr metric. $\mathcal{CH}^+$ is a (smoothly extendible) Cauchy horizon emanating from timelike infinity i^+, and $\mathcal{T}$ a timelike ring singularity.

with the Kerr metric (see [12], Section 6). In [15,16], he discovered that the Kerr Cauchy horizon suffers from a blue-shift instability (we will return to this in Section 4). This instability mechanism gave credence to his original formulation of Strong Cosmic Censorship, which precisely aimed at avoiding locally naked singularities. It is also important to note that the very existence of the Kerr metric already demonstrates that Cauchy horizons (and the breakdown of determinism they cause) cannot be completely ruled out from General Relativity, which mandates a *genericity condition* in the statement of Strong Cosmic Censorship (see Section 3), excluding the Kerr metric (see the related discussion on blue-shift instability in Section 4).

1.4. *The modern formulation of Strong Cosmic Censorship in terms of global hyperbolicity*

Fifty years after its inception by Penrose, the statement of Strong Cosmic Censorship remains true to its original spirit, while benefiting from the modern developments in the theory of the initial value problem for the Einstein equations. In 1974, Penrose already proposed a connection between Strong Cosmic Censorship and "Leray's condition of global hyperbolicity" [9]. This insight later allowed for a clear formulation of Strong Cosmic Censorship in terms of the maximal globally hyperbolic development (MGHD), a concept introduced in the work of Choquet-Bruhat and Geroch [17]. Given a suitable[3] initial hypersurface (viewed as a snapshot of spacetime at time zero), its MGHD represents the maximal spacetime that can be predicted by means of the Einstein equations from the data posed on this initial hypersurface (see Section 2.3). A failure of determinism, however, arises if the MGHD is *extendible as a solution to the Einstein equations*, for such an extension would be, by definition, not predicted by the initial conditions generating the MGHD. In light of these connections, Geroch and Horowitz [10] introduced the first modern formulation of Strong Cosmic Censorship as the impossibility of extending the MGHD for generic asymptotically flat[4] initial data. Such an approach *excludes de facto timelike singularities* since they are incompatible with global hyperbolicity. Cauchy horizons, should they exist, are part of the future boundary of the MGHD and, as such, the sole threat to Laplacian determinism

[3]By suitable initial hypersurface, it is meant a triplet (M, g_0, K_0), where (M, g_0) a complete three-dimensional Riemannian manifold, and K_0 a symmetric tensor such that (M, g_0, K_0) satisfy the Einstein constraint equations.

[4]Asymptotic flatness is a natural assumption when studying gravitational collapse, our main focus in this review. In the context of Cosmology, it is also interesting to study Strong Cosmic Censorship for compact initial data, see Section 7.

by potentially allowing for spacetime extendibility. The MGHD viewpoint leads to various similar formulations of Strong Cosmic Censorship [11,18–21], summarized in the following conjecture.

Conjecture 1.1 (Strong Cosmic Censorship, modern formulation). *For generic, asymptotically flat, complete, and regular initial data, the Maximal Globally Hyperbolic Development (MGHD) is inextendible as a solution to the Einstein equations.*

The precise definitions of "generic", "regular", and "solutions" are left intentionally vague to allow flexibility in solving the conjecture. It turns out that the validity of the conjecture varies according to the precise meaning of a "solution" to the Einstein equations (and its regularity): physically, this is interpreted as the fact that determinism only holds for spacetime metrics belonging to an adequate regularity class [22–24]; we will come back to this in Section 2.3.

1.5. *The modern Weak and Strong Cosmic Censorship as two independent statements*

While the statement of Weak Cosmic Censorship, i.e., the absence of naked singularities in generic gravitational collapse (see [10,11]), has remained essentially unchanged since its 1969 inception [4], the Strong Cosmic Censorship statement from Conjecture 1.1 no longer explicitly refers to locally naked singularities contrary to Penrose's original version [9]. As such, the modern approach to both Weak and Strong Cosmic Censorship renders both statements logically independent; in other words, *Weak Cosmic Censorship does not follow from the resolution of Strong Cosmic Censorship*[5]. We note, however, that the presence of locally naked singularities inside black holes (irrelevant to Weak Cosmic Censorship in the form the conjecture is currently stated [10,11]) remains an obstruction to Strong Cosmic Censorship, both in its original formulation [5,9] and its modern form adopted in Conjecture 1.1 (we will discuss this in more depth in Section 5).

1.6. *Strong Cosmic Censorship as a mathematical problem*

Beyond the fundamental issue of determinism posed by Strong Cosmic Censorship, Conjecture 1.1 turns out to be a source of profound mathematical questions at the crossroads of PDEs and Differential Geometry. A foundational problem in evolutionary PDEs is well-posedness: does there exist a unique solution to the equations with prescribed initial data (existence/uniqueness)? Of course, in the context of the Einstein equations, one could interpret the existence of a unique MGHD as settling this question. However, it is desirable to instead interpret the extendibility of the MGHD as *a failure of global uniqueness for the Einstein equations*—an idea first formulated by Dafermos [21,25,26]—for there is no relativistic principle providing a canonical extension of the MGHD—should the MGHD be extendible: therefore, all possible extensions are arbitrary.

We remark that, as abstractly formulated in Conjecture 1.1, the ontology of Strong Cosmic Censorship is unrelated, prima facie, to singularities and black holes. However, it turns out that an important obstruction to Strong Cosmic Censorship precisely originates in the interior of black holes, as shown by the example of the Kerr metric that we now reexamine under a modern light. In PDE terminology, the MGHD of Kerr initial data terminates at a Cauchy horizon, which is future-extendible and, thus, would represent a counter-example to Conjecture 1.1 had we not included the word "generic" in its formulation (see Section 3.3). It is precisely *because* the MGHD

[5] In Christodoulou's words "'strong' and 'weak' in reference to the two cosmic censorship conjectures is not a happy one, for it seems to suggest that the 'strong' conjecture implies the 'weak' one, which is not the case" [11].

of the Kerr metric is singularity-free (hence extendible) that, ironically, determinism now appears in jeopardy. Therefore, a restorationist strategy for determinism should involve showing that for *generic perturbations* of the Kerr black hole, the MGHD features a singular terminal boundary, across which no (future) extension is possible: this is indeed the hope entertained in the early works on blue-shift instability of Penrose [15] and Penrose–Simpson [16] that we discussed above. More generally, a resolution of Conjecture 1.1 would require showing the presence of a singularity in the interior of a generic black hole and thus answering the following fundamental problem:

Problem 1.2. *In the context of generic gravitational collapse, is the terminal boundary of a black hole interior singular and, if so, how strong is the singularity?*

The strength of the singularity is crucial, moreover, in that it precludes a failure of determinism in its corresponding regularity class [23,27,28]. The theme of singularities in the context of black holes (and more broadly within general relativity) is indeed intimately connected to Cosmic Censorship; we will discuss this, together with the celebrated singularity theorems in Section 2.

Taking a step back, we may also connect such a mathematical endeavor with the more general theme of *dynamical singularity formation in evolutionary PDEs*, a class of differential equations to which the Einstein equations belong. Finally, we emphasize that the connection between a singularity within the system of PDEs and the geometric inextendibility of the MGHD is subtle due to the covariant nature of the Einstein equations[6], which gives rise to fascinating geometric problems (see, in particular, Sections 3 and 4).

We conclude this section emphasizing that Strong Cosmic Censorship, despite its concise statement in Conjecture 1.1, encompasses a web of mathematical problems and provides a testing field to rigorously understand deep physical concepts such as singularities or determinism. The conjecture has nowadays become a field of research on its own and the source of many exciting mathematical developments. Progress toward its resolution is poised to have repercussions that extend far beyond applications in General Relativity, particularly within the field of PDEs.

2. Initial value problem and global hyperbolicity

2.1. *Laplacian determinism from Newtonian mechanics to general relativity*

Newtonian mechanics has long been employed in the well-known classical theory of gravitation, the predecessor of General Relativity. In Newtonian mechanics, the motion of a point-particle of mass m in a gravity field $\vec{g}$ is governed by Newton's second law giving rise to the following ODE:

$$\begin{aligned} m\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} &= m\vec{g}(x(t), y(t), z(t), t), \\ \vec{x}(t=0) = (x_0, y_0, z_0), \quad \frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(t=0) &= (\dot{x}_0, \dot{y}_0, \dot{z}_0). \end{aligned} \tag{2.1}$$

Equation (2.1) is par excellence deterministic in the sense of Laplace: indeed, for prescribed initial positions (x_0, y_0, z_0) and velocities $(\dot{x}_0, \dot{y}_0, \dot{z}_0)$, (2.1) has a unique (possibly blowing-up in finite-time) solution. Hence, the initial state of the test particle fully determines its future evolution[7] Analogously in General Relativity, the motion of a freely falling test-particle (also

[6]The diffeomorphism-invariance of the theory indeed requires demonstrating the presence of a singularity in a coordinate-independent fashion, which is a challenging task in general.

[7]For N particles, however, (2.1) becomes a system of nonlinear ODEs and its solutions may blow-up in finite time (e.g. if two particles collide), which can cause a breakdown of determinism for this model if $N \geq 3$, see [29], Chapter III.

known as a freely-falling *observer*) on a Lorentzian manifold $(\mathscr{M}, g)$ is governed by the timelike geodesic equation

$$\begin{gathered}\frac{\mathrm{d}^2\vec{x}^\alpha}{\mathrm{d}\tau^2} + \Gamma^\alpha_{\sigma\mu}(g)\frac{\mathrm{d}\vec{x}^\sigma}{\mathrm{d}\tau}\frac{\mathrm{d}\vec{x}^\mu}{\mathrm{d}\tau} = \vec{0},\\ \vec{x}(\tau=0) = (x_0, y_0, z_0, t_0), \quad \frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(\tau=0) = (\dot{x}_0, \dot{y}_0, \dot{z}_0, \dot{t}_0),\\ g\left(\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(\tau=0), \frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(\tau=0)\right) = -1.\end{gathered} \tag{2.2}$$

where $\tau \geq 0$ represents the observer's proper time (note that $\vec{x}$ is now a four-dimensional spacetime vector) and $\Gamma^\alpha_{\sigma\mu}(g)$ are the Christoffel symbols of the spacetime metric g (see e.g. [6,7]).

Remark. A timelike geodesic represents the wordline of a freely falling massive test-particle. The analogue for a massless test-particle (e.g. a photon) is provided by the null geodesic equation

$$\begin{gathered}\frac{\mathrm{d}^2\vec{x}^\alpha}{\mathrm{d}\tau^2} + \Gamma^\alpha_{\sigma\mu}(g)\frac{\mathrm{d}\vec{x}^\sigma}{\mathrm{d}\tau}\frac{\mathrm{d}\vec{x}^\mu}{\mathrm{d}\tau} = \vec{0},\\ \vec{x}(\tau=0) = (x_0, y_0, z_0, t_0), \quad \frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(\tau=0) = (\dot{x}_0, \dot{y}_0, \dot{z}_0, \dot{t}_0),\\ g\left(\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(\tau=0), \frac{\mathrm{d}\vec{x}}{\mathrm{d}t}(\tau=0)\right) = 0.\end{gathered} \tag{2.3}$$

Together, (2.2) and (2.3) provide the equation of a so-called causal geodesic.

It is possible for a causal geodesic to exit the spacetime $(\mathscr{M}, g)$ in finite proper time τ_{exit} (geodesic incompleteness, see Section 2.4). In this situation, Laplacian determinism begs the question: do the equations predict what happens to this observer? We distinguish two possibilities:

(A) The observer encounters a terminal singularity, in the sense that (2.2) cannot be continued beyond the exit proper time τ_{exit}. This situation occurs for instance at the $\{r = 0\}$ singularity in the interior of the Schwarzschild black hole (see Section 3 for further details).

(B) The observer does not encounter a terminal singularity, in the sense that $(\mathscr{M}, g)$ can be (future) extended as a Lorentzian manifold in which (2.2) makes sense for $\tau > \tau_{\text{exit}}$. This situation occurs at the Cauchy horizon in the interior of the Kerr black hole, see Figure 1.

A modern interpretation of Scenario (A) is that Laplacian determinism is respected in this case [3,22,23,30], highlighting the significance of singularities for Strong Cosmic Censorship. Scenario (B), however, is not entirely conclusive; indeed, unveiling potential violations of determinism in this situation additionally requires examining the predictability issues associated with equations governing the spacetime metric (e.g., Einstein equations), see Section 2.3. We note that the situation of Scenario (B) underscores the primacy of *spacetime extendibility* in resolving Conjecture 1.1, suggesting it is a more fundamental concept than the continuation of individual worldlines solutions of (2.2) or (2.3), as we discuss in Section 2.3.

2.2. *Cauchy surfaces and global hyperbolicity*

We now return to formulating the previously-described failure of determinism in terms of spacetime extendibility. We first describe general Lorentzian-geometric notions that do not require the Einstein equations (2.4) to be satisfied, starting with the concept of a Cauchy surface.

Definition 2.1 (Cauchy surface). *A hypersurface Σ is a Cauchy surface for the Lorentzian manifold $(\mathscr{M}, g)$ if it is met exactly once by any inextendible timelike curve.*

The concept of a Cauchy surface is crucial in making sense of determinism within General Relativity for it provides an initial time hypersurface on which to prescribe initial conditions for the Einstein equations (2.4) (we will come back to this in Section 2.3). A spacetime in which any event is causally connected to such a Cauchy surface is called *globally hyperbolic*.

Definition 2.2 (Globally hyperbolic spacetime). *The Lorentzian manifold* $(\mathscr{M}, g)$ *is called globally hyperbolic if it admits a Cauchy surface* Σ.

Note the region $\{t \geq 0\}$ on the maximal analytically-extended Schwarzschild spacetime (see Section 3.2) is globally hyperbolic, while the region $\{t \geq 0\}$ on the maximal analytically extended Kerr spacetime is not [6,31]. Not every spacetime $(\mathscr{M}, g)$ is globally hyperbolic, however, it is always possible [7] to construct a globally hyperbolic subset of $(\mathscr{M}, g)$ by considering the (interior of the) domain of dependence $\mathscr{D}(\Sigma)$ of an achronal hypersurface Σ: in the case of $\Sigma = \{t = 0\}$ in the Kerr spacetime, $\mathscr{D}(\Sigma)$ is the subregion of $\{t \geq 0\}$ below the Cauchy horizon, see Figure 1.

2.3. *The Einstein equations of spacetime dynamics and their initial value problem*

The Einstein field equations. Contrary to Newtonian mechanics, the ontology of General Relativity is not about the trajectory of particles per sé, but instead about the dynamics of *spacetime*, modeled by a four-dimensional Lorentzian metric $(\mathscr{M}, g)$ governed by the Einstein equations

$$\mathrm{Ric}_{\mu\nu}(g) - \tfrac{1}{2}\mathrm{Scal}(g) g_{\mu\nu} = 8\pi \mathbb{T}_{\mu\nu}, \tag{2.4}$$

where $\mathrm{Ric}_{\mu\nu}(g)$ is the Ricci curvature tensor of the spacetime metric g and $\mathrm{Scal}(g)$ its scalar curvature. Here, $\mathbb{T}_{\mu\nu}$ is the stress-energy tensor of the matter in presence, with $\mathbb{T}_{\mu\nu} \equiv 0$ for the Einstein vacuum equations.

The Energy conditions. The positivity of energy is reflected by the so-called energy conditions [6,31], the weakest of which being the null energy condition for solutions of (2.4):

$$\mathbb{T}(v, v) \geq 0 \quad \text{for any null spacetime vector } v. \tag{2.5}$$

Note that (2.5) can equivalently be formulated purely in terms of the spacetime metric g as $\mathrm{Ric}(v, v) \geq 0$ for any null spacetime vector v, independently of (2.4). Analogously, one defines the strong energy condition (which indeed implies the validity of the weak energy condition (2.5)):

$$\mathrm{Ric}(T, T) \geq 0 \quad \text{for any timelike spacetime vector } T. \tag{2.6}$$

Classical solutions versus weak solutions. A classical solution g of (2.4) must possess twice differentiable components $g_{\mu\nu}$ in some appropriate coordinate system (dubbed a C^2 solution). Interestingly, however, it is possible to make sense of weak solutions of (2.4) for continuous metrics g with locally square-integrable Christoffel symbols [24] (dubbed a H^1_{loc} solution). These notions play an important role in the formulation of Strong Cosmic Censorship, see below.

Initial value problem and maximal globally hyperbolic development. We are now ready to view the Einstein equations (2.4) as an initial value problem analogously to (2.1). We call the following a suitable initial data set for the Einstein equations (2.4) (say in vacuum, i.e., with $\mathbb{T} = 0$).

(1) A complete three-dimensional regular Riemmanian manifold (Σ, g_0), which will later be viewed as the first fundamental form of the spacetime metric g with respect to Σ.
(2) A symmetric regular two-tensor K_0, which will later be viewed as the second fundamental form of the spacetime metric g with respect to Σ.
(3) Assume (g_0, K_0) satisfy the Einstein constraints equations induced by (2.4) (see e.g. [32]).

To solve the initial value problem, one needs to treat Σ as a Cauchy surface and construct a globally hyperbolic spacetime solution of (2.4) consisting of the wordlines of timelike curves starting at Σ: the resulting Lorentzian manifold $(\mathcal{M}, g)$ is the Globally Hyperbolic Development (GHD) of the initial data (Σ, g_0, K_0), more precisely:

Definition 2.3. *A globally hyperbolic development (GHD) $(\mathcal{M}, g)$ of suitable initial data (Σ, g_0, K_0) is a globally hyperbolic Lorentzian manifold $(\mathcal{M}, g)$ satisfying (2.4) as a classical solution together with an embedding $i : \Sigma \to \mathcal{M}$ such that $i^*(g) = g_0$, $i^*(K) = K_0$, where K is the second fundamental form of $i(\Sigma)$ in $\mathcal{M}$ with respect to the future normal and $i(\Sigma)$ is a Cauchy surface in $(\mathcal{M}, g)$.*

The cornerstone of the initial value problem for the Einstein equations is given by the following result of Choquet-Bruhat and Geroch [17] stating that there exists a unique maximal GHD of the prescribed initial data set (Σ, g_0, K_0), considered to be *the* solution of (2.4) for such initial data.

Theorem 2.4 ([17,33]). *Given a suitable initial data set (Σ, g_0, K_0), there exists a GHD $(\mathcal{M}, g)$ that is an extension of any other GHD of the same initial data. This GHD $(\mathcal{M}, g)$ is unique up to isometry and is called the maximal globally hyperbolic development (MGHD) of the initial data set (Σ, g_0, K_0).*

Remark. In the above Theorem 2.4, we assumed regular initial data; however, rougher solutions of (2.4) are of considerable mathematical and physical interest. The state of the art is given by the Klainerman–Rodnianski–Szeftel bounded-L^2 curvature theorem [34] proving the existence of a weak solution of (2.4) in vacuum for initial data with two square-integrable derivatives (i.e., H^2_{loc} in some coordinates). It is possible, however, to solve (2.4) for even rougher initial data, provided they possess a special structure, for instance in the case of impulsive gravitational waves [35–39].

Returning to Laplacian determinism: inextendibility of the MGHD. The MGHD $(\mathcal{M}, g)$ constructed in Theorem 2.4 consists of worldlines of inextendible timelike curves (i.e., observers). But what if the MGHD $(\mathcal{M}, g)$ itself was extendible, i.e., there exists a Lorentzian manifold $(\bar{\mathcal{M}}, \bar{g})$ solution of (2.4) and a proper embedding $J : \mathcal{M} \to \bar{\mathcal{M}}$? Due to the hyperbolic nature of (2.4), one *cannot* expect such an extension to be unique [22,26], which already demonstrates a failure of global uniqueness for the initial value problem described above. This failure is also accompanied by a breakdown of Laplacian determinism, in the sense that the timelike curves which were inextendible in $(\mathcal{M}, g)$ become extendible within $(\bar{\mathcal{M}}, \bar{g})$: since the extension $(\bar{\mathcal{M}}, \bar{g})$ is arbitrary, so is the fate of these observers, which is not uniquely determined by their initial conditions at the Cauchy surface Σ. The validity of Laplacian determinism thus mandates the **inextendibility of the MGHD as a solution to the Einstein equations**, which motivated our formulation of Strong Cosmic Censorship as Conjecture 1.1 in Section 1.4. To conclude this section, we will formulate three physically-relevant notions of MGHD inextendibility of decreasing stength:

(I) (C^0 Strong Cosmic Censorship) The MGHD is inextendible as a C^0 Lorentzian manifold. This is the strongest possible inextendibility requirement at present, which is a pointwise failure of continuity/boundedness of the metric coefficients in *every coordinate system*.

(II) (H^1 Strong Cosmic Censorship) The MGHD is inextendible as a C^0 Lorentzian manifold with square-integrable Christoffel symbols. Inextendibility at the H^1 regularity precludes any extension of the MGHD even as a weak solution to the Einstein equations (2.4).

(III) (C^2 Strong Cosmic Censorship) The MGHD is inextendible as a C^2 Lorentzian manifold with square-integrable Christoffel symbols. Inextendibility at the C^2 regularity precludes any extension of the MGHD as a classical solution to the Einstein equations (2.4).

Remark. It might be possible, in principle, to prove directly that the MGHD is inextendible as a (weak) solution to (2.4). Instead, however, all known works on this question prove inextendibility as a Lorentzian manifold within a given regularity class such as the ones of Statement (I), (II), or (III).

2.4. *Singularity theorems and the problem of geodesic incompleteness*

If the Strong Cosmic Censorship of Conjecture 1.1 is true, then what is the mechanism leading to the inextendibility of the MGHD? In most cases, MGHD inextendibility is caused by the presence[8] of a **singularity** on its the terminal boundary. The celebrated Penrose singularity theorem [41] shows that the presence of a (closed) trapped surface, namely a topological sphere whose outgoing null mean curvature is negative (see e.g. [42]), leads to geodesic incompleteness.

Theorem 2.5 ([41]). *If $(\mathcal{M}, g)$ admits a non-compact Cauchy hypersurface, contains a closed trapped surface and satisfies the energy condition (2.5), then it is future-causally geodesically incomplete.*

Note that the singularity inside Schwarzschild's black hole (which, of course, contains a trapped surface) is indeed associated to causal geodesic incompleteness (see Section 3.2). Because a surface being trapped is a stable condition with respect to small perturbations, Theorem 2.5 shows that geodesic incompleteness (interpreted as a form of singularity) "is a robust prediction of the general theory of relativity" [43]. However, it turns out that Theorem 2.5 merely asserts the existence of incomplete causal geodesics, which is not necessarily associated to a blow-up of curvature (or other gauge-invariant metric quantities): this is indeed what happens for the MGHD of the Kerr spacetime, when incomplete geodesics reach its Cauchy horizon as depicted in Figure 1 *despite the absence of any singularity at the Cauchy horizon* (Scenario (B) in Section 2.1). For this reason, Theorem 2.5 is sometimes called the Penrose incompleteness theorem [23,30].

As a conclusion to this section, we emphasize that Conjecture 1.1 requires to go beyond the statement provided by Theorem 2.5 and specify the nature of the geodesic incompleteness, particularly whether it is associated to a metric singularity strong enough to preclude extendibility in the sense of statements (I), (II) or (III) described in the previous Section 2.3. Because trapped surfaces are the precursor of black holes, this mathematical endeavor is naturally connected to the fundamental Open Problem 1.2, which will require addressing in any resolution of Conjecture 1.1.

3. Extendibility and singularity structure of particular solutions

While the extendibility properties of specific spacetimes technically has no bearing on Conjecture 1.1, since it only concerns *generic solutions* of the Einstein equations (2.4), it is instructive to examine well-known examples in the light of the concepts developed in the earlier Section 2.

3.1. *Minkowski spacetime and other geodesically complete spacetimes*

The Minkowski spacetime $(\mathbb{R}^{3+1}, m)$ is the flat solution to (2.4) in a vacuum given by

$$m = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{3.1}$$

$(\mathbb{R}^{3+1}, m)$ is well-known to be causally geodesically complete, and stable to small perturbations from the celebrated work of Christodoulou–Klainerman [44] and Lindblad–Rodnianski [45]. Since both the Minkowski spacetime and its pertubations are singularity-free and causally geodesically complete, how do they fit in within the inextendibility requirements of Conjecture 1.1? As it turns out, any timelike geodesically complete spacetime (including $(\mathbb{R}^{3+1}, m)$) satisfies the most demanding notion of Strong Cosmic Censorship, i.e., Statement (I) in Section 2.3.

[8] Except if the MGHD is timelike geodesically complete, e.g. for Minkoski spacetime, in which case it is C^0 inextendible [40] despite the absence of singularities, see Section 3.1 for a detailed discussion.

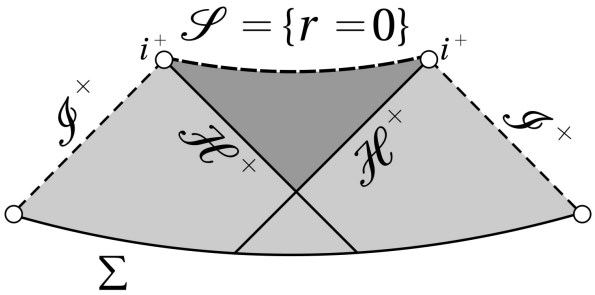


Figure 2. Schwarzschild as the MGHD of a two-ended asymptotically flat hypersurface Σ.

Theorem 3.1 ([40]). *A smooth (at least C^2) time-oriented Lorentzian manifold that is time-like geodesically complete and globally hyperbolic is C^0-inextendible.*

Therefore, Theorem 3.1 shows that in proving Conjecture 1.1, one can now restrict our attention to timelike geodesically incomplete spacetimes, a large class of which is provided by spacetimes with a trapped surface (including any sub-extremal black holes), by Theorem 2.5.

3.2. *The Schwarzschild black hole*

The Schwarzshild spacetime $(\mathbb{R}^{3+1}, g_S)$ is a one-parameter family of black hole solutions to (2.4) in a vacuum indexed by their mass parameter $M > 0$, which takes for following form when $r \neq 2M$:

$$g_S = -\left(1 - \frac{2M}{r}\right) \mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1} \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\varphi^2). \tag{3.2}$$

The spacetime geometry of the Schwarzshild spacetime $(\mathbb{R}^{3+1}, g_S)$ is depicted in Figure 2, with the event[9] horizon $\mathscr{H}^+ = \{r = 2M\}$ and the spacelike singularity $\mathscr{S} = \{r = 0\}$. The $\{r = 0\}$ singularity is known to be terminal, in the sense that the Kretschmann scalar is infinite [46] (curvature blow-up): it is thus well-known that the Schwarzschild spacetime is C^2-inextendible [42] and, therefore, satisfies Strong Cosmic Censorship in the sense of Statement (III) in Section 2.3. Moreover, it is also classically known [31,46] that in-falling observers reach $\mathscr{S} = \{r = 0\}$ in finite proper time and experience *infinite tidal deformations*, which correspond, at least heuristically, to a metric blow-up on the C^0 level. These considerations beg the following question [42]: is the Schwarzschild spacetime inextendible in a stronger sense, e.g. as in Statement (I) of Section 2.3? The affirmative answer to this longstanding question was obtained by Sbierski [47] in the following theorem:

Theorem 3.2 ([47]). *The MGHD of the (two-ended asymptotically flat) hypersurface $\{t = 0\}$ of the Schwarzschild spacetime g_S is C^0-inextendible (to the future).*

3.3. *The Reissner–Nordström and Kerr black holes*

The Schwarschild spacetime (3.2) represents the simplest black hole, uncharged and non-rotating. Its charged analogue is the Reissner–Nordström spacetime $(\mathbb{R}^{3+1}, g_{\mathrm{RN}})$ of charge e for $0 < |e| \leq M$, which takes the following form when $r \neq r_+(M, e)$, where $r_\pm(M, e) = M \pm \sqrt{M^2 - e^2} > 0$:

$$g_{\mathrm{RN}} = -\left(1 - \frac{2M}{r} + \frac{e^2}{r^2}\right) \mathrm{d}t^2 + \left(1 - \frac{2M}{r} + \frac{e^2}{r^2}\right)^{-1} \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\varphi^2) \tag{3.3}$$

[9] The event horizon $\mathscr{H}^+ = \{r = 2M\}$ is a regular hypersurface for g_S under the appropriate choice of coordinates [6,31].

and is a solution of the Einstein–Maxwell equations:

$$\mathrm{Ric}_{\mu\nu}(g) - \tfrac{1}{2}\mathrm{Scal}(g)g_{\mu\nu} = 8\pi\underbrace{\left(g^{\alpha\beta}F_{\alpha\nu}F_{\beta\mu} - \tfrac{1}{4}F^{\alpha\beta}F_{\alpha\beta}g_{\mu\nu}\right)}_{=\mathbb{T}^{\mathrm{EM}}_{\mu\nu}}, \quad \nabla^{\mu}F_{\mu\nu} = 0. \tag{3.4}$$

Note that $r_{\pm}(M,e)$ correspond to the roots of $(1-(2M/r)+e^2/r^2)$, with $\mathscr{H}^+ = \{r = r_+(M,e)\}$ being the event horizon and $\mathscr{CH}^+ = \{r = r_-(M,e)\}$ the Cauchy horizon: the MGHD of the Cauchy surface $\{t=0\}$ is $\{t \geq 0, r < r_-(M,e)\}$ and the Penrose diagram of the Reissner–Nordström spacetime is the same as that of Figure 1. The extremal Reissner–Nordström spacetime corresponds to the case $|e| = M$. The rotating analogue of the Schwarzschild spacetime is the celebrated Kerr [48] black hole of angular momentum $0 < |a| \leq M$ of the following form when $r \neq r_+(M,a)$:

$$\begin{aligned} g_{\mathrm{K}} = &-\frac{1-\frac{2M}{r}+\frac{a^2}{r^2}}{1+\frac{a^2\cos^2(\theta)}{r^2}}(\mathrm{d}t - a\sin^2(\theta)\,\mathrm{d}\varphi)^2 + \frac{1+\frac{a^2\cos^2(\theta)}{r^2}}{1-\frac{2M}{r}+\frac{a^2}{r^2}}\mathrm{d}r^2 \\ &+[r^2+a^2\cos^2(\theta)]\mathrm{d}\theta^2 + \frac{\sin^2(\theta)}{r^2+a^2\cos^2(\theta)}[a\,\mathrm{d}t - (r^2+a^2)\,\mathrm{d}\varphi]^2, \end{aligned} \tag{3.5}$$

and is a solution of (2.4) in vacuum ($\mathbb{T} = 0$). Analogously, $\mathscr{H}^+ = \{r = r_+(M,a)\}$ is the event horizon and $\mathscr{CH}^+ = \{r = r_-(M,a)\}$ the Cauchy horizon, where $r_{\pm}(M,a) = M \pm \sqrt{M^2-a^2}$; the MGHD is $\{t \geq 0, r < r_-(M,a)\}$ and the Penrose diagram of the Kerr spacetime is given in Figure 1. Finally, the extremal Kerr spacetime corresponds to $|a| = M$.

As already discussed in Section 1, the MGHD of the Reissner–Nordström and Kerr spacetimes are *smoothly and non-uniquely extendible* across the Cauchy horizon $\mathscr{CH}^+$ and, therefore, do not even satisfy the least demanding inextendibility of Statement (III). This smooth extendibility makes for a serious threat to Laplacian determinism, all the more that the Kerr black hole is often considered to be a realistic model for astrophysical black holes. However, we emphasize that the Reissner–Nordström/Kerr spacetimes are not counter-examples to Strong Cosmic Censorship as formulated in Conjecture 1.1, precisely because they are *not generic solutions* of (2.4), in view of their stationary character. To prove Conjecture 1.1, one needs in particular prove the following:

Problem 3.3. *Show that for generic and smooth small perturbations of the Reissner–Nordström/Kerr initial data, the MGHD solving (2.4) is inextendible as a suitably regular Lorentzian manifold.*

This genericity condition may appear as an all-too-convenient fiat in order to restore determinism. However, we emphasize that, in the Penrose blue-shift scenario [4,15], the Reissner–Nordström/Kerr interior region (i.e., the region $\{r < r_+\}$ beneath the event horizon) is conjecturally *unstable to small perturbations*. Since a physically realistic solutions must be stable to small perturbations, the Reissner–Nordström/Kerr interior region and their smooth extendibility can be deemed to be unphysical and therefore, it is legitimate to exclude them from the picture of gravitational collapse. That said, it is yet unclear to which extent Penrose's blue-shift instability operates as a restorationist of determinism; we will discuss this in more detail in Section 4.

3.4. *The Oppenheimer–Snyder spacetime*

The first description of gravitational collapse came with the celebrated Oppenheimer–Snyder [49] black hole solution of (2.4) with dust (a pressure-less perfect fluid with velocity u^{α}, density ρ)

$$\mathbb{T}^{\mathrm{dust}}_{\alpha\beta} = \rho u_{\alpha} u_{\beta}. \tag{3.6}$$

It consists of a ball of homogeneous dust of finite radius, outside of which the spacetime coincides with the Schwarzschild (3.2) metric, and terminates at a spacelike singularity $\mathscr{S} = \{r = 0\}$ as

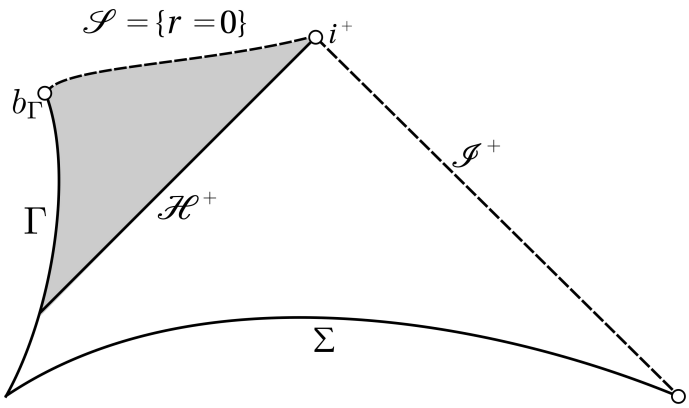


Figure 3. Penrose diagram of Oppenheimer–Snyder [49] spacetime (gravitational collapse).

depicted in Figure 3. In view of Theorem 3.2, we may conjecture that the Oppenheimer–Snyder spacetime is also C^0-inextendible. However, we emphasize that the Oppenheimer–Snyder spacetime does not represent a generic[10] instance of gravitational collapse, due to its high number of symmetries and the homogeneity of the dust ball, see Section 5 for further discussions.

3.5. *The naked and locally naked singularities*

In the seminal work [51], Christodoulou constructed examples of naked singularities solutions of the Einstein-scalar-field equations, i.e., (2.4) with the following scalar field stress-energy tensor

$$\mathbb{T}^{\mathrm{SF}}_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \tfrac{1}{2}(g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi)g_{\mu\nu}, \quad \Box_g\phi = 0, \tag{3.7}$$

and moreover arising as a model of gravitational collapse (see Section 5). The Penrose diagram of Christodoulou's spacetime is depicted on the left of Figure 4, and b_Γ is a naked singularity. Analogously [51] also constructs a *locally naked singularity* b_Γ inside a black hole region, see the right panel of Figure 4: in both cases, a Cauchy horizon $\mathscr{CH}_\Gamma$ emanates from the singular point b_Γ and across it, the spacetime is extendible[11] as a solution to (2.4), (3.7). The Christodoulou spacetimes do not have smooth initial data (see the footnote below) and therefore, do not, strictly speaking, pose a threat to Conjecture 1.1, which is formulated for solutions of (2.4) with smooth initial data. The same is true for the recently-constructed naked singularity solutions of the vacuum (2.4) in the breakthrough of Shlapentokh-Rothman–Rodnianski [54]. Within spherically symmetric matter models, naked singularities with *smooth initial data*, however, have been observed numerically [55,56]. They pose a threat to Strong Cosmic Censorship, similarly to the Cauchy horizons discussed in Section 3.3. In Section 5, we will discuss strategies to show that such (locally) naked singularities are non-generic solutions, in order to prove Conjecture 1.1.

[10] In fact, even within the spherically-symmetric solutions of the Einstein-dust equations, the Oppenheimer–Snyder solution is non-generic, and some of its perturbations actually feature a naked singularity [50]! The standard interpretation is that dust is an unphysical matter model in these circumstances in which the fluid density becomes large [13,42].

[11] as a solution of the Einstein-scalar-field equations in the (rough) BV class defined by Christodoulou in [52]. In fact, the extension is $C^{1,\theta}$, which has the same regularity as the initial data on a suitable Cauchy surface [53].

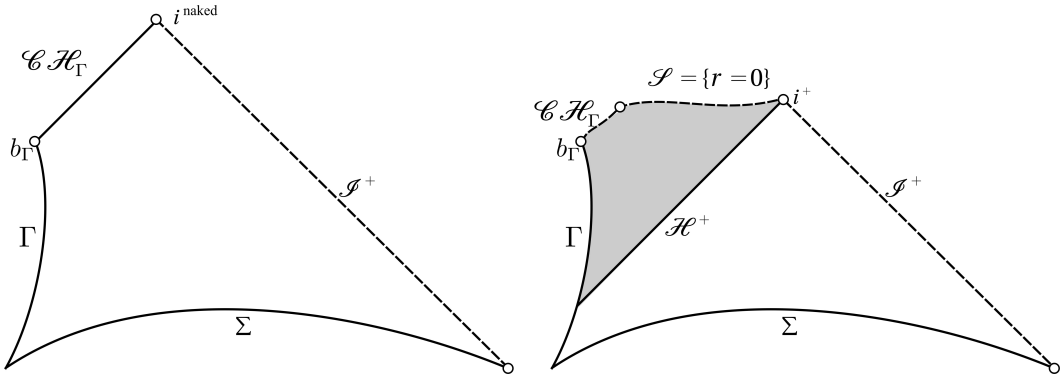


Figure 4. Left: Penrose diagram of a naked singularity spacetime. Right: Penrose diagram of a spacetime with a locally naked singularity inside a black hole.

4. Cauchy horizons in dynamical black holes and blue-shift instability

4.1. *Blue-shift instability: heuristics and numerics*

The celebrated blue-shift instability mechanism at the Reissner–Nordström/Kerr Cauchy horizon is the cornerstone of Strong Cosmic Censorship as already envisioned by Penrose [4].

Consider two freely-falling observers: an in-falling timelike incomplete geodesic A entering the black hole region, reaching the Cauchy horizon $\mathscr{C}\mathscr{H}^+$ in finite proper time, and a complete timelike geodesic B with infinite proper time remaining outside of the black hole forever. The geometry depicted in Figure 5 is such that a pulse sent by B towards A with constant frequency becomes infinitely blue-shifted as A reaches $\mathscr{C}\mathscr{H}^+$. The numerical study of linear fields on a fixed black hole (see Section 4.2) confirms the blue-shift instability scenario, see [15,57]. How this blue-shift mechanism impacts the Cauchy horizon stability for solutions of the Einstein equations (2.4) (so-called back-reaction problem), however, remains to be seen, and will be discussed below.

4.2. *Blue-shift in linear test-field theories on a fixed stationary black hole*

Before turning to the full nonlinear dynamics of the Einstein equations (2.4), it is useful to first consider the behavior of linear test fields on a fixed sub-extremal Reissner–Nordström (i.e., (3.3) with $0 < |e| < M$) or Kerr (i.e., (3.5) with $0 < |a| < M$) black hole, starting with the wave equation[12]

$$\Box_g \phi = 0, \quad \text{with } g = g_{\text{RN}} \text{ or } g = g_{\text{K}}. \tag{4.1}$$

The first theoretical approach to establishing blow-up at the Cauchy horizon for solutions of (4.1) (and their spin-weighted analogues, including electromagnetic and linearized gravitational fields) is due to McNamara [59] using a scattering method. Initial data are posed to be trivial at the past event horizon $\mathscr{H}^-$ and a fast-decaying radiation field at past null infinity $\mathscr{I}^-$. The key idea, clarified in [58], is to construct a *transmission map* taking the initial data on $\mathscr{I}^-$ to the Cauchy horizon $\mathscr{C}\mathscr{H}^+$, and use the t-invariance of (3.3) (or (3.5)) to show that the local energy of the transmitted solution ϕ blows up at the Cauchy horizon $\mathscr{C}\mathscr{H}^+$; in other words, $\phi \notin H^1_{\text{loc}}$, which is a linear analogue of the formulation (II) of Strong Cosmic Censorship from Section 2.3. The rigorous construction of the transmission map and its non-trivial character were proved in the work of Dafermos–Shlapentokh-Rothman [58] establishing the following linear instability result[13]:

[12] We recall that (4.1) is often used as a linear toy model for the Einstein equations, due to their hyperbolic nature.

[13] Anterior work of Sbierski [60] on Gaussian beams also provides solutions with arbitrarily large local energy at $\mathscr{C}\mathscr{H}^+$.

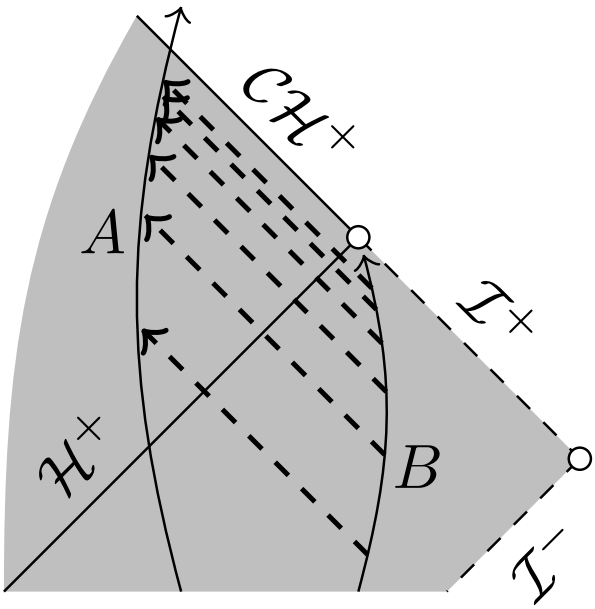


Figure 5. Blue-shift effect at the Reissner–Nordström/Kerr Cauchy horizon, figure from [58].

Theorem 4.1 ([58]). *There exists a solution ϕ to (4.1) such that $r\phi_{|\mathcal{H}^-} \equiv 0$ and $r\phi_{|\mathcal{I}^-}$ decays at an arbitrarily fast polynomial rate, yet the local energy of ϕ is infinite at the Cauchy horizon $\mathcal{CH}^+$.*

How to reconcile the existence of merely one solution with infinite local energy at the Cauchy horizon with the genericity requirement of Strong Cosmic Censorship? A well-known argument (see e.g. [61]) shows that for (4.1), the existence of one blowing-up solution gives rise to blow-up for all initial data (in the same regularity class) except a co-dimension one subset: blow-up is thus generic in this sense. However, scattering-type blow-up arguments à la McNamara do not provide a constructive condition to decide whether a given solution to (4.1) blows up at $\mathcal{CH}^+$ or not. A different approach to Cauchy horizon blow-up for (4.1), which relies on propagating inverse-polynomial lower bounds in time, has been proposed by Luk–Oh [61] on the (sub-extremal) Reissner–Nordström interior region (3.3), then by Luk–Sbierski [62] on the (sub-extremal) Kerr interior (3.5), starting from characteristic initial data on the event horizon $\mathcal{H}^+$. See also the more recent linear results for (4.1) by Ma–Zheng [63] and Luk–Oh–Shlapentokh-Rothman [64].

Theorem 4.2 ([61–64]). *If ϕ is a solution of (4.1) is such that its energy of along the event horizon $\mathcal{H}^+$ obeys some polynomial upper and lower bounds then its local energy is infinite at $\mathcal{CH}^+$.*

We also mention the recent work of Sbierski [65] establishing an analogous result to Theorem 4.2 for the equations of linearized gravity; such results are poised to be used in an eventual proof of Strong Cosmic Censorship for the full (nonlinear) Einstein equations. We finally emphasize that the assumptions of Theorem 4.2 can be shown to be satisfied for *generic* (sufficiently regular and localized) initial data at the Cauchy surface $\{t = 0\}$ for (4.1), which becomes an asymptotic-in-time problem in the black hole exterior; we will return to this in Section 6.

4.3. *Nonlinear dynamics and spherically-symmetric matter models*

It is natural to consider Strong Cosmic Censorship for solutions to the Einstein equations in vacuum, as we will discuss below. In order to understand delicate nonlinear issues like the

ones associated to Strong Cosmic Censorship, it is useful to first look at spherically symmetric models, which are simpler to track. However, the well-known Birkhoff theorem (see e.g. [6,31]) shows that *all spherically symmetric solutions of (2.4) in vacuum are trivial*, in the sense that they are isometric to the Schwarzchild (3.2) or Minkowski metric (3.1). In other words: the spherically-symmetric Einstein equations do not have a dynamical nature in vacuum and are merely constraints. Therefore, to study spherically symmetric dynamical problems in General Relativity, one must consider (2.4) coupled to a non-trivial matter model $\mathbb{T}$. Ideally, the matter model should reintroduce the wave degrees of freedom lost by imposing spherical symmetry to the Einstein equations [11], which is what inspired Christodoulou to study the Einstein-scalar-field model (2.4), (3.7) in spherical symmetry in his series of works [51,53,66]. We will give a non-exhaustive lists of such scalar-field matter models, in increasing order of sophistication.

- **(Einstein-scalar-field)** The model (2.4), (3.7) was studied by Christodoulou in [52,53,66] in his work on gravitational collapse already mentioned in Section 3.5. Note that the Schwarzschild metric (3.2) is a solution of (2.4), (3.7) with $\phi \equiv 0$. The Einstein-scalar-field spherically symmetric model does not allow for the presence of Cauchy horizons emanating from timelike infinity i^+, because of the absence of angular momentum or electric charge, so we will not discuss it further in this section (see, however, Section 5).
- **(Einstein–Maxwell-scalar-field)** To study Cauchy horizons while remaining in spherical symmetry, Dafermos [25,67] considered (2.4) coupled with the stress-energy tensor

$$\mathbb{T}_{\mu\nu} = \mathbb{T}^{\mathrm{EM}}_{\mu\nu} + \mathbb{T}^{\mathrm{SF}}_{\mu\nu}, \quad \Box_g \phi = 0, \quad \nabla^\mu F_{\mu\nu} = 0 \tag{4.2}$$

(recall $\mathbb{T}^{\mathrm{EM}}_{\mu\nu}$ is defined in (3.4)). The Reissner–Nordström metric (3.3) is a solution of (2.4), (4.2) with $\phi \equiv 0$. Note that the uncharged scalar field ϕ is not directly coupled to the Maxwell field $F_{\mu\nu}$. In the absence of charged matter, $F_{\mu\nu}$ is subjected to a spherically-symmetric rigidity analogous to Birkhoff theorem [13,25,67] and takes the trivial form

$$F = \frac{e}{r^2}\Omega^2 \mathrm{d}u \wedge \mathrm{d}v, \quad \text{where } g = -\Omega^2 \mathrm{d}u\mathrm{d}v + r^2 \mathrm{d}\sigma_{\mathbb{S}^2}, \tag{4.3}$$

and $e \neq 0$ is a constant. Importantly, and contrary to Christodoulou's electromagnetism-free model, it is not possible to study gravitational collapse spacetimes within this model because such a situation requires the presence of a center $\Gamma = \{r = 0\}$ (see Section 5) where both the spacetime metric g and the electromagnetic field F are regular. The rigid form (4.3) and the $e \neq 0$ are obviously incompatible with such a requirement at $r = 0$.
- **(Einstein–Maxwell-charged-scalar-field)** To remedy the Maxwell field rigidity of (4.3), one can study a *charged scalar field* [13,68], namely the coupled stress-energy tensor:

$$\begin{gathered} \mathbb{T}_{\mu\nu} = \mathbb{T}^{\mathrm{EM}}_{\mu\nu} + \mathbb{T}^{\mathrm{CSF}}_{\mu\nu}, \quad \mathbb{T}^{\mathrm{CSF}}_{\mu\nu} = \Re(D_\mu\phi\overline{D_\nu\phi}) - \tfrac{1}{2}\left(g^{\alpha\beta}D_\alpha\phi\overline{D_\beta\phi}\right)g_{\mu\nu}, \\ g^{\mu\nu}D_\mu D_\nu\phi = 0, \quad \nabla^\mu F_{\mu\nu} = \mathrm{i}q_0\Im(\bar{\phi}D_\nu\phi), \quad D_\mu = \nabla_\mu + \mathrm{i}q_0 A_\mu, \quad F = \mathrm{d}A, \end{gathered} \tag{4.4}$$

where ∇_μ is the Levi-Civita connection of the spacetime metric g, $q_0 \neq 0$ is a coupling constant, and A_μ is the electromagnetic potential. Because the electromagnetic form F is now coupled to the charged scalar field ϕ, it is no longer subjected to the rigidity of (4.3). Therefore, (4.4) is suitable as a spherically-symmetric model of gravitational collapse.

All the above spherically-symmetric models are generally considered as toy models for the Einstein vacuum equations (i.e., (2.4) with $\mathbb{T} = 0$). We emphasize that, strictly speaking, spherically-symmetric solutions of (2.4) are non-generic and thus have no bearing to Conjecture 1.1; nonetheless, one can formulate an analogous simplified version of Conjecture 1.1 within the class of spherically symmetric solution.

Finally, to complicate matters further, note that it is possible to study the *massive Klein–Gordon* variant of all these models (see already Section 6.3), replacing $\Box_g\phi = 0$ by

$$\Box_g\phi = m^2\phi, \tag{4.5}$$

where $m^2 > 0$. The Einstein–Klein–Gordon equation, in contrast, is often viewed as a matter model in its own sake (as opposed to a toy model for vacuum dynamics) which features different phenomena than the Einstein vacuum equations [28,69–72].

4.4. *Einstein dynamics: nonlinear stability of the Cauchy horizon*

One of the key issues in resolving Strong Cosmic Censorship is to understand the consequences of the blue-shift instability for dynamical perturbations of the Reissner–Nordström and Kerr spacetimes solving the Einstein equations (2.4) either in vacuum or for a reasonable matter model. In view of the examples of Section 3 (among the only examples known at the time of Penrose, see e.g. [6]), it might seem reasonable to speculate that the blue-shift instability "destroys the Cauchy horizon", and causes the dynamical appearance of a terminal singularity, the only known example of which being $\mathscr{S} = \{r = 0\}$ in the Schwarzschild black hole (3.2). In the context of the vacuum Einstein equations, Penrose [5] writes "there are some reasons for believing that generic perturbations away from spherical symmetry will not change the spacelike of the singularity". Interpreting this scenario in the context of Conjecture 1.1, one is tempted to speculate about the validity Strong Cosmic Censorship in the C^0 sense, i.e., Statement (I) from Section 2.3 for small perturbations of the Kerr black hole, which leads to the following conjecture.

Conjecture 4.3 (Spacelike singularity conjecture). *The MGHD of small perturbations of the Kerr metric (3.5) solving the Einstein equations (2.4) in a vacuum terminates at a spacelike singularity $\mathscr{S}$ and has the same Penrose diagram as Figure 2. Moreover, the MGHD is C^0-future inextendible.*

Dafermos studied this problem in spherical symmetry for solutions of the Einstein–Maxwell-scalar-field system (4.2) converging to a sub-extremal Reissner–Nordström black hole, and *surprisingly* proved that the Cauchy horizon ends up being stable to perturbations [25,67], blue-shift instability notwithstanding! As a consequence, he also showed that the spacetime metric is continuously extendible, thus falsifying the spherically-symmetric version of Conjecture 4.3 and the C^0 version of Strong Cosmic Censorship given in Statement (I) in Section 2.3. The author later [68] established the Cauchy horizon stability for the more general Einstein–Maxwell-charged-scalar-field system (4.4) under which the convergence to a sub-extremal Reissner–Nordström black hole is slower (see below), which necessitates the use of different methods from [25,67]. Below and in the rest of this review, we use the word *"Cauchy horizon" for a null component of the MGHD terminal boundary foliated by topological spheres of non-zero volume* (see [13,14,22]).

Theorem 4.4. *Consider spherically symmetric characteristic initial data on $\underline{C}_{in} \cup \mathscr{H}^+$ as depicted in Figure 6, and under the assumption that (g,ϕ) converge to a sub-extremal Reissner–Nordström black hole with inverse-polynomial upper bounds for ϕ on the event horizon $\mathscr{H}^+$.*

- *[25,67,73] For the Einstein–Maxwell-scalar-field system (4.2), the solution admits a non-empty Cauchy horizon $\mathscr{CH}^+$ across which the metric g is continuously extendible.*
- *[68] For the Einstein–Maxwell-charged-scalar-field system (4.4), the solution admits a non-empty Cauchy horizon $\mathscr{CH}^+$.*

The question of continuous extendibility for the Einstein–Maxwell-charged-scalar-field system (4.4) is left open in Theorem 4.4; it is directly related to the pointwise boundedness of the scalar field at the Cauchy horizon, i.e., is it true that $\phi \in L^\infty$? Because the decay assumptions on the event horizon $\mathscr{H}^+$ are weaker than in the uncharged-scalar-field case, this issue becomes

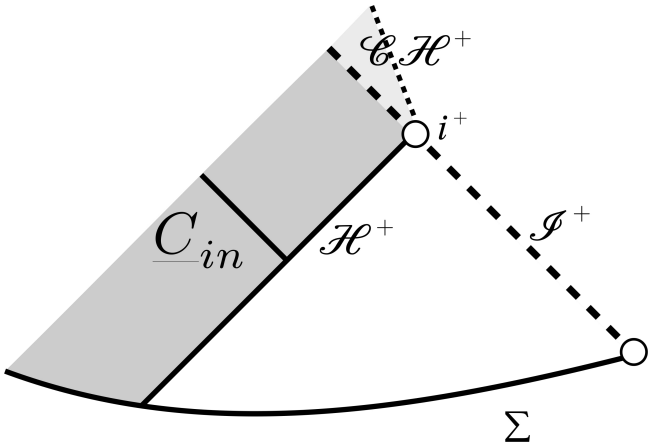


Figure 6. Characteristic initial data in the black hole interior and continuous extension across the Cauchy horizon $\mathscr{C}\mathscr{H}^+$ for Theorems 4.4, 4.5 and 4.6.

delicate and depends on the exact profile of the characteristic initial data $\phi_{|\mathscr{H}^+}$ (see already Theorem 4.5 below). More precisely, the decay assumptions in Theorem 4.4 take the following form in an Eddington–Finkelstein advanced-time coordinate v on $\mathscr{H}^+$ as $v \to +\infty$, for some $s > 3/4$:

$$[1+|v|^s]\phi_{|\mathscr{H}^+}(v) \text{ and its derivative are bounded in an unweighted-in-}v\text{ norm.} \tag{4.6}$$

For the Einstein–Maxwell-scalar-field system (4.2), one can prove (4.6) for $s = 3$ for spherically-symmetric solutions with regular and localized initial data [74], while for the Einstein–Maxwell-charged-scalar-field (4.4), the conjecture [28] is that for black holes of asymptotic charge $e \neq 0$

$$s(q_0 e) = 1 + \sqrt{1-4(q_0 e)^2} \quad \text{if } |q_0 e| < \tfrac{1}{2}, \quad s(q_0 e) = 1 \text{ otherwise.} \tag{4.7}$$

(we will discuss these decay rates in Section 6). In Theorem 4.4, continuous extendibility can be proven[14] provided (4.6) with $s > 1$ is assumed, which only makes sense for (4.2) or (4.4) if $|q_0 e| < 1/2$. However, under the assumptions of Theorem 4.4, some solutions are C^0-extendible, but others are not [28], depending on the scalar field oscillations on $\mathscr{H}^+$: the author proved with Kehle that

Theorem 4.5 ([28]). *Under the same assumptions as Theorem 4.4 for the Einstein–Maxwell-charged-scalar-field (4.4), assume additionally that $\phi_{|\mathscr{H}^+}$ oscillates sufficiently. Then $\phi \in L^\infty$ and (g,ϕ) are continuously extendible across the Cauchy horizon $\mathscr{C}\mathscr{H}^+$.*

The oscillation condition of Theorem 4.5 involves the finiteness of integrals of the form

$$\left|\int_1^\infty \phi_{|\mathscr{H}^+}(v)\mathrm{e}^{-\mathrm{i}[\omega_{\mathrm{res}}+o(1)]v}\mathrm{d}v\right| < \infty, \tag{4.8}$$

where $\omega_{\mathrm{res}} = q_0 e[1/r_+ - 1/r_-] \in \mathbb{R}$; (4.8) is a semi-integrability condition in particular satisfied if (4.6) holds for $s > 1$, but not necessarily otherwise. The oscillation condition (4.8) is conjectured to be satisfied [75,76] for generic solutions of (2.4), (4.4), see the discussion in Section 6.

We now return to the vacuum Einstein equations (2.4) and discuss the breakthrough of Dafermos–Luk [22] proving that vacuum black holes converging to Kerr (3.5) also admit a C^0-extendible Cauchy horizon, and thus offering a definitive disproof of Conjecture 4.3.

[14] For $s > 1$, continuous extendibility is deduced from a stronger extendibility property of (g,ϕ) in a $W^{1,1}$-type norm.

Theorem 4.6 ([22]). *Consider characteristic initial data on $\underline{C}_{in} \cup \mathcal{H}^+$ as depicted in Figure 6, and under the assumption that g converge to a sub-extremal Kerr black hole with inverse-polynomial upper bounds on the event horizon $\mathcal{H}^+$. Then, the resulting solution $(\mathcal{M}, g)$ of (2.4) in vacuum admits a non-empty Cauchy horizon $\mathcal{CH}^+$ across which the metric g is continuously extendible.*

Together with the stability of the Kerr black hole exterior proven in [77,78] (see Section 6), Theorem 4.6 also falsifies the C^0-version of Strong Cosmic Censorship (statement (I) in Section 2.3). However, despite Conjecture 4.3 being false, it is still possible that during gravitational collapse, *part of the terminal boundary is spacelike*; this will a major topic discussed in Section 5.

4.5. *Einstein dynamics: nonlinear instability of the Cauchy horizon and mass inflation*

We now return to Open Problem 3.3, which is an essential piece in a potential resolution of Strong Cosmic Censorship. The Reissner–Nordström/Kerr Cauchy horizons dynamical stability obtained in Section 4.4 begs the question: what are the consequences of the blue-shift instability discussed in Section 4.2 in the nonlinear dynamics picture? The falsification of Conjecture 4.3 already shows that the blue-shift instability *does not destroy the Cauchy horizon*, which remains continuously extendible. Could it then be that it is also smoothly extendible as in the Reissner–Nordström/Kerr cases discussed in Section 3? The key insight comes from the mass inflation scenario of Poisson–Israel [79,80] and Ori [81] (see also their precursor [82]). They study (2.4) in spherical symmetry coupled to a rudimentary matter model in which two clouds of dust are transported in the ingoing and outgoing null directions. Their analysis involves the Hawking mass, defined in terms of the area-radius function r of the spherically-symmetric metric g as

$$\rho := \frac{r[1 - g(\nabla r, \nabla r)]}{2}. \tag{4.9}$$

They showed that, for non-trivial ingoing and outgoing event-horizon radiation falling-off at an inverse polynomial rate as in the theorems of Section 4.4, the Hawking mass ρ and the Riemann curvature blow up at the Cauchy horizon. In summary, Strong Cosmic Censorship, at least in the weaker sense of Statement (II) and Statement (III) from Section 2.3, can still be valid *despite the existence of Cauchy horizon* in dynamical spacetimes. For the Einstein–Maxwell-scalar-field system (4.2), Dafermos established the blow-up of ρ under the inverse-polynomial decay assumption (4.6) together with an associated pointwise lower bound assumption [25,67]. The blow-up of curvature was established by Luk–Oh [73] under a more relaxed inverse-polynomial lower bound assumption on the energy of $\phi_{|\mathcal{H}^+}$, while the blow-up of ρ was obtained under similar assumptions in [64]. Curvature blow-up for the charged scalar field case (4.4) was obtained in the author's work [68,83]. All these results are summarized in the following theorem.

Theorem 4.7. *In the setting and under the same assumptions as Theorem 4.4, assume additionally that (4.6) also holds as a lower bound on the event horizon $\mathcal{H}^+$ in a suitable norm.*

- *[25,64,67,73] For the Einstein–Maxwell-scalar-field system (4.2), the Cauchy horizon $\mathcal{CH}^+$ is weakly singular, in the sense that ρ and certain curvature components blow-up.*
- *[68,83] For the Einstein–Maxwell-charged-scalar-field system (4.4), the Cauchy horizon $\mathcal{CH}^+$ is weakly singular, in the sense that certain curvature components blow-up.*

We emphasize that lower bounds on the scalar field are essential in Theorem 4.7, since when $\phi \equiv 0$ (Reissner–Nordström case), the Cauchy horizon $\mathcal{CH}^+$ is not singular (recalling Section 3). In Section 6, we will see that these lower bounds on the event horizon $\mathcal{H}^+$ are satisfied for spherically symmetric solutions of (4.2) with generic, localized and regular Cauchy data in the black hole exterior [84], thus connecting with Conjecture 1.1 restricted to spherical symmetry.

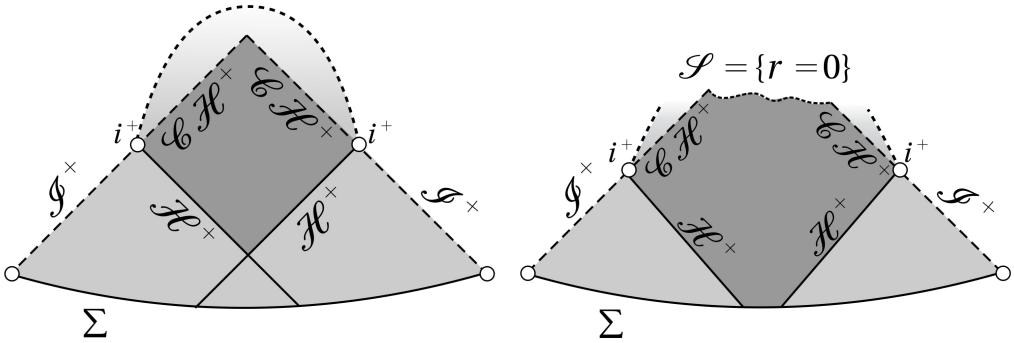


Figure 7. Left: two-ended black hole with a Cauchy horizon but no spacelike singularity. Right: two-ended black hole with a spacelike singularity $\mathscr{S}$ and a Cauchy horizon $\mathscr{CH}^+$. In both cases, the spacetime can be continuously-extended through $\mathscr{CH}^+$.

In terms of inextendibility, Theorem 4.7 leads to the proof that there *does not exist* any C^2 extension of the metric [68,73,83] across $\mathscr{CH}^+$, consistently with Statement (III) in Section 2.3.

Finally, we return to the Einstein vacuum equations ((2.4) with $\mathbb{T} = 0$) for which the first examples of a black hole with a weakly singular Cauchy horizons were constructed by Luk [85].

Theorem 4.8 ([85]). *There exists a class of black hole interior solutions of the Einstein vacuum equation without any symmetry assumptions featuring a weakly singular Cauchy horizon $\mathscr{CH}^+$.*

The stability estimates in Theorem 4.6 are consistent with Theorem 4.8 and suggest that a generic perturbation of the Kerr black hole features a weak singularity similar to the constructions of [85]. The analogous result to Theorem 4.7 for (2.4) in vacuum, however, is still open at present.

4.6. *Strong Cosmic Censorship for two-ended spacetimes*

The results in Section 4.4 are stated in terms of local regions of spacetime as depicted in Figure 6; however, Conjecture 1.1 is stated in terms of global inextendibility. In Penrose's original formulation [4], Strong Cosmic Censorship is in the context of gravitational collapse. As we will see in Section 5 when discussing the global structure of collapse, this involves a one-ended spacetime with a regular center Γ. However, outside of the study of gravitational collapse, it is also possible to consider the global structure of two-ended spacetimes, which also corresponds to that of the Reissner–Nordström (3.3) or Kerr (3.5) spacetime (see Figure 1). The first global spherically-symmetric result for (4.2) is due to Dafermos [21] who proved that for *small perturbations* of Reissner–Nordström metric, there is no spacelike singularity at all and only a null Cauchy horizon $\mathscr{CH}^+$ as depicted in the left panel of Figure 7. However, it follows from [86] that there exists *large perturbations* of the Reissner–Nordström metric obeying the assumptions of Theorem 4.4 featuring a non-empty spacelike singularity $\mathscr{S} = \{r = 0\}$, as depicted on the right panel of Figure 7 (see Section 4.7). For (2.4) in vacuum, Dafermos–Luk [22] considered small perturbations of the Kerr metric and also showed the absence of spacelike singularities in this case. These results are summarized in the following theorem.

Theorem 4.9. *(Two-ended dynamical black holes with no spacelike singularity).*

- *[21,87] Consider a bifurcate two-ended event horizon $\mathscr{H}^+$ on which the decay assumptions of Theorem 4.4 are satisfied, and assume additionally that (the characteristic data induced by) (g,ϕ) on $\mathscr{H}^+$ is a small perturbation of $(g_{\mathrm{RN}},0)$ in an appropriate norm. Then*

the solution to (4.2) (or (4.4)) has no spacelike singularity and its terminal boundary only consists of the (bifurcate) Cauchy horizon $\mathscr{C}\mathscr{H}^+$, as depicted in the left of Figure 7.

- *[22] Consider a bifurcate two-ended event horizon $\mathscr{H}^+$ on which the decay assumptions of Theorem 4.6 are satisfied, and assume additionally that (the characteristic data induced by) g on $\mathscr{H}^+$ is a small perturbation of g_K in an appropriate norm. Then the solution to (2.4) in a vacuum has no spacelike singularity and its terminal boundary only consists of the (bifurcate) Cauchy horizon $\mathscr{C}\mathscr{H}^+$, as depicted in the left of Figure 7.*

Returning to the issue of Strong Cosmic Censorship, we finally state the only known definitive statement of inextendibility for generic spherically symmetric two-ended solutions of (4.2).

Theorem 4.10 (C^2 Strong Cosmic Censorship in spherical symmetry [73,84]). *There exists an open and dense subset (for the topology induced by a smooth norm) within the set of admissible spherically-symmetric initial data such that their MGHD for (4.2) is C^2 (future) inextendible.*

The above result by Luk–Oh [73,84] can be interpreted as a full resolution of a spherically-symmetric version of Conjecture 1.1 at the C^2 level (Statement (III) of Section 2.3) for two-ended initial data. We emphasize that their C^2 extendibility result applies to both cases in Figure 7. Furthemore, we note that a recent result of Sbierski [88] showing that the spacetimes of Theorem 4.9 are also C^1 inextendible (which is an intermediate version of Strong Cosmic Censorship between Statement (II) and Statement (III) in Section 2.3). Finally, the Hawking mass blow-up (conditionally) obtained in Theorem 4.7 leads to conjecture the validity of a H^1 version Strong Cosmic Censorship [26], i.e., Statement (II) from Section 2.3, still open at this time.

4.7. *Two-ended black holes with a spacelike singularity and a Cauchy horizon*

Interior of charged hairy black holes. The author's works [86] and [89] with Li considered the interior of black holes with a scalar hair– namely a non-trivial stationary scalar field on the event horizon. These spacetimes are spherically-symmetric solutions of the Einstein–Maxwell-scalar-field equations (4.2) in [86] and Einstein–Maxwell-charged-scalar-field (4.4) in [89] and their terminal boundary is entirely-spacelike, thus they do not possess any Cauchy horizon (compare with Theorem 4.4). They have a two-ended $\mathbb{R}\times\mathbb{S}^2$ spatial topology, but are not associated to an asymptotically flat[15] black hole exterior, and thus do not play any direct role in Strong Cosmic Censorship as formulated in Conjecture 1.1. They will be used, however, in the asymptotically flat black hole constructions discussed in the paragraph below. For now, we state the results of [86,89].

Theorem 4.11 ([86,89]). *There exists spatially-homogeneous solutions of the Einstein–Maxwell-(un)charged-scalar-field equations (4.2) or (4.4) with a constant scalar field data $\phi\equiv\epsilon\neq 0$ on a two-ended event horizon $\mathscr{H}_L\cup\mathscr{H}_R$, whose MGHD terminates at a spacelike boundary $\mathscr{S}=\{r=0\}$.*

(1) *(violent collapse, [86]) For the Einstein–Maxwell-scalar-field equations (4.2), $\mathscr{S}=\{r=0\}$ is approximated by a Kasner metric of exponents $(1-2C\epsilon^2, C\epsilon^2, C\epsilon^2)$ for a constant $C>0$.*

(2) *(fluctuating collapse, [89]) For the Einstein–Maxwell-charged-scalar-field equations (4.4), $\mathscr{S}=\{r=0\}$ is approximated by a Kasner metric of positive exponents $(1-2p(\epsilon), p(\epsilon), p(\epsilon))$, where $p(\epsilon)\in(0,1/2)$ is an algebraic function of a highly-oscillating quantity $\alpha(\epsilon)$ of the form*

$$\alpha(\epsilon)\approx C\sin(\omega_0\epsilon^{-2}+O(\log(\epsilon^{-1})))\quad \text{as } \epsilon\to 0$$

for $C\neq 0$ and $\omega_0\in\mathbb{R}-\{0\}$. Moreover, for some values of ϵ, a Kasner bounce takes place.

[15] However, Zheng recently [90] constructed a class of asymptotically AdS (the analogue of asymptotic flatness in the presence of a negative cosmological constant) black holes to which Theorem 4.11 applies, see the discussion in Section 7.

To clarify the terminology, we recall the Kasner spacetime [91], a solution of the Einstein vacuum equations. For $(p_1, p_2, p_3) \in \mathbb{R}^3$ such that $p_1 + p_2 + p_3 = 1$ (the Kasner exponents), define:

$$g_{\mathrm{K}} = -\mathrm{d}t^2 + t^{2p_1}\mathrm{d}x_1^2 + t^{2p_2}\mathrm{d}x_2^2 + t^{2p_3}\mathrm{d}x_3^2. \tag{4.10}$$

If $(p_1, p_2, p_3) = (1, 0, 0)$ (or perturbations), g_{K} is flat (i.e., locally isometric to the Minkowski metric (3.1)) but otherwise, $\mathscr{S} = \{t = 0\}$ is a spacelike singularity. In Theorem 4.11, the Kasner exponents p_i are all positive (which is consistent with the known stability results in this case [92], see Section 7). Note, however, that in the uncharged-scalar-field case (4.2), the Kasner exponents degenerate to the Minkowskian endpoint $(p_1, p_2, p_3) = (1, 0, 0)$ when ϵ becomes arbitrarily small.

The terminology "Kasner bounce" originates from the foundational works of BKL [93,94] on Kasner dynamics for the Einstein equations. These celebrated heuristics proposed an instability mechanism for Kasner metrics (4.10) with one negative p_i leading to the transition ("Kasner epoch") to a stable Kasner metric with all positive p_i's. Statement (2) of Theorem 4.11 provides the first rigorous quantitative study of the long-hypothesized Kasner bounce phenomenon.

Construction of asymptotically flat two-ended black holes. A gluing argument[16] allows to construct two-ended asymptotically flat black holes satisfying the assumptions of Theorem 4.4, but whose terminal boundary features a spacelike component $\mathscr{S}'$ isometric to a subset of $\mathscr{S} = \{r = 0\}$ in Theorem 4.11 (see [89][Section 1.2] for details). In particular, one can in principle construct spherically-symmetric asymptotically flat solutions of Einstein–Maxwell-charged-scalar-field equations (4.4) exhibiting the Kasner bounce described in the previous paragraph.

In the case of the Einstein–Maxwell-charged-scalar-field equations (4.4), however, one would additionally require decay estimates in the black hole exterior to finalize the construction (see the discussion in Section 6). For the Einstein–Maxwell-scalar-field equations (4.2) case, these estimates have been obtained in [74,84] and thus the gluing argument provides a large class of two-ended asymptotically flat black holes with the rightmost Penrose diagram of Figure 7, where part of the singularity $\mathscr{S} = \{r = 0\}$ is spacelike and described by the Kasner-like asymptotics of Theorem 4.11.

We note that for such constructions, Theorem 4.11 only provides estimates near $\mathscr{S}$ away from the intersection $\mathscr{CH}^+ \cap \mathscr{S}$. However, the results of [95] apply to this situation and provide a quantitative description of the null-spacelike transition region near $\mathscr{CH}^+ \cap \mathscr{S}$, see Section 5.4.

5. Gravitational collapse and (locally) naked singularities

5.1. *Mathematical setting of gravitational collapse*

Gravitational collapse is the widely-accepted theory accounting for the existence of astrophysical black holes, as pioneered in the celebrated work of Oppenheimer–Snyder [49] (recall the discussion in Section 3.4). In this astrophysical process, a star whose initial state is free of any trapped surface develops a trapped surface at a later time. The mathematical setting for gravitational collapse imposes[17] an initial surface Σ that is diffeomorphic to $\mathbb{R}^3$ and free of any trapped surface, with the origin of $\mathbb{R}^3$ corresponding to the center of the star Γ. Of course, not all such spacetimes undergo a gravitational collapse process (e.g. recall from Section 3.1 that perturbations of the Minkowski spacetime do not admit any trapped surface). Nonetheless, in Definition 5.1 below, we introduce spacetimes that are initially free of any trapped surface: thus gravitational collapse

[16] While the gluing argument of [89] only provides two-ended asymptotically-flat black holes featuring a piece of hairy black hole singularity $\mathscr{S}$, we conjecture that one-ended versions of such construction are also possible. Unlike their two-ended counterparts, one-ended black holes play a major role in gravitational collapse, see Section 5.

[17] Topological censorship considerations regarding an initial hypersurface Σ indeed indicate that Σ must have a trivial topology (like $\mathbb{R}^3$ in Definition 5.1) if it is free of any trapped surface, see [96] for interesting results in this direction.

may take place if the conditions leading to the formation of a trapped surface[18] are satisfied at a later time. For simplicity, we restrict ourselves to the vacuum case ((2.4) with $\mathbb{T} = 0$), keeping in mind that Definition 5.1 generalizes in the presence of a suitable matter model (see e.g. [13]).

Definition 5.1. *We say a Lorentzian manifold $(\mathcal{M}, g)$ solution of (2.4) is a spacetime suitable for gravitational collapse if it is the MGHD of suitable initial data set (Σ, g_0, K_0) such that*

- *Σ is diffeomorphic to $\mathbb{R}^3$ and free of any trapped surface.*
- *(g_0, K_0) are asymptotically flat, i.e., tends to the Minkowski solution $(m, 0)$ as $r \to +\infty$.*

The central axis Γ is defined as the wordline of the timelike geodesic starting at the (diffeomorphic image of the) origin in $\mathbb{R}^3$ in the future normal direction to Σ.

Note that a spacetime suitable for gravitational collapse possesses one asymptotically flat end ("one-ended spacetime"), as opposed to the two-ended spacetimes (including the Reissner–Nordström (3.3) and Kerr (3.5) MGHDs) discussed in Section 4.6 whose topology is that of $\mathbb{R} \times \mathbb{S}^2$.

5.2. *From Weak Cosmic Censorship to Strong Cosmic Censorship*

In the celebrated series [11,51–53,66], Christodoulou studied a model of spherical gravitational collapse in considering spherically symmetric solutions of (3.7) satisfying the conditions of Definition 5.1. As we discussed in Section 4.3, such spacetimes cannot feature a Cauchy horizon emanating from i^+, but Christodoulou constructed [51] examples of (locally) naked singularity solutions with a Cauchy horizon $\mathcal{CH}_\Gamma$ emanating from the center Γ, as depicted in Figure 4, with initial data in a Hölder class. Christodolou specifically studied a rougher class of spherically symmetric solutions of (3.7) within the BV (bounded variation) class in the series [51–53,66]. He proved that (locally) naked singularities satisfying the condition $\rho/r \not\to 0$ (recall the definition in (4.9)) on their past light cone are non-generic, in the sense they have finite co-dimension within this class. Combining the results of [51–53,66] with a sharp extension criterion given in [11,53] that implies any naked singularity satisfies the condition $\rho/r \not\to 0$, results in the following theorem.

Theorem 5.2 ([11,51–53,66]). *For spherically-symmetric initial data belonging to the BV class, the MGHD solution of (3.7) satisfy one the following:*

(a) *are geodesically complete, with the same Penrose diagram as the Minkowski spacetime.*
(b) *or possess a black hole region with a spacelike terminal boundary $\mathcal{S} = \{r = 0\}$ with the Penrose diagram of Figure 3. In this case, the MGHD is C^2 (future) inextendible.*
(c) *or feature a (locally) naked singularity b_Γ with the Penrose diagram of Figure 4.*

Moreover, for generic initial data within the above class, only options (a) or (b) are possible.

As is clear from Theorem 5.2, the only obstruction to determinism for the above model are the Cauchy horizons $\mathcal{CH}_\Gamma$ emanating from the center Γ which are, in principle, extendible. In view of Theorem 5.2 showing that spacetimes with $\mathcal{CH}_\Gamma \neq \emptyset$ are non-generic within the (spherically-symmetric) BV class, one can also interpret Theorem 5.2 as validating a version of Strong Cosmic Censorship at the C^2 level (Statement (III) from Section 2.3) for rough and spherically-symmetric initial data. We emphasize that, as in Section 4, Cauchy horizons constitute the main obstruction to Strong Cosmic Censorship. However, contrary to the Cauchy horizon $\mathcal{CH}^+$ emanating of i^+ discussed in Section 4 that is subjected to a blue-shift instability, there is no such mechanism for Cauchy horizons $\mathcal{CH}_\Gamma$ emanating of Γ and therefore, no hope for $\mathcal{CH}_\Gamma$ to be singular. That is why, should Strong Cosmic Censorship be true, it must be that $\mathcal{CH}_\Gamma = \emptyset$ for generic solutions.

[18] There is a rich mathematical theory for the formation of trapped surface, see e.g. [24,66,97–101].

Remark. The strategy employed by Christodoulou [52,53] to obtain $\mathscr{CH}_\Gamma = \emptyset$ for generic solutions is to perturb an hypothetical solution with $\mathscr{CH}_\Gamma \neq \emptyset$ in order to obtain a sequence of trapped surfaces converging to b_Γ: it is then easy to see that $\mathscr{CH}_\Gamma = \emptyset$ for the perturbed solution [13].

While Theorem 5.2 only provides C^2 inextendibility, the spacelike nature of the terminal MGHD boundary $\mathscr{S} = \{r = 0\}$ (and Theorem 3.2) leads to conjecturing that a generic spherically-symmetric solution is C^0 inextendible. This expectation is reinforced by the fact that $\mathscr{S} = \{r = 0\}$ is asymptotically Schwarzschild towards i^+ as proven in [102] (see also [103,104]). For the more sophisticated Einstein–Maxwell-charged-scalar-field model (4.4) which additionally allows to study Cauchy horizons emanating from i^+ (as we explained in Section 4.3), the global picture of spherically-symmetric gravitational collapse is more subtle and several aspects will be discussed in Sections 5.3 and 5.4. We emphasize, however, that the Christodoulou argument showing that $\mathscr{CH}_\Gamma = \emptyset$ generically is local near Γ, and thus, independent of the dynamics near i^+. We may thus speculate [13,87] that $\mathscr{CH}_\Gamma = \emptyset$ for generic spherically-symmetric solutions of the Einstein–Maxwell-charged-scalar-field model (4.4) (see [97,105] for recent results in this direction).

5.3. *The breakdown of weak null singularities*

As we saw in Section 4 (Theorem 4.4), the MGHD terminal boundary is not *everywhere-spacelike* for spherically-symmetric solutions of the Einstein–Maxwell-charged-scalar-field model (4.4) (at least, under decay conditions of the form (4.6) on the event horizon that are conjectured to be satisfied for regular localized solutions). The falsification of Conjecture 4.3 for this model ironically begs the question: *does there exist a spacelike singularity at all*? The two-ended spacetimes of Section 4.6 provide examples in which $\mathscr{S} = \emptyset$ and the Cauchy horizon $\mathscr{CH}^+$ "closes-off" the spacetime, as depicted in Figure 7, but, as explained in Section 5.1, such spacetimes are not models of gravitational collapse and do not satisfy the assumptions of Definition 5.1. For one-ended spacetime suitable for gravitational collapse, it is possible, in principle, that $\mathscr{CH}^+$ the Cauchy horizon emanating from i^+ is the only component of the MGHD terminal boundary and thus meets the central axis Γ at its endpoint b_Γ as depicted in Figure 8. In fact, an example of a spacetime with the Penrose diagram of Figure 8 has been constructed in [106] using a gluing procedure. However, the author proved that, providing the Cauchy horizon $\mathscr{CH}^+$ is weakly singular, it cannot "close-off" the spacetime, and therefore, the Penrose diagram of Figure 8 is impossible [87].

Theorem 5.3 (Breakdown of weak null singularities[87]). *Let a spacetime suitable for gravitational collapse in the sense of Definition 5.1 with a future terminal boundary $\mathscr{B}$ including a Cauchy horizon $\mathscr{CH}^+ \subset \mathscr{B}$ emanating from i^+. Assume that $\mathscr{CH}^+$ is weakly singular, in the sense of Theorem 4.7. Then $\mathscr{CH}^+ \underset{\neq}{\subset} \mathscr{B}$, i.e., the Penrose diagram of the spacetime is not given by Figure 8.*

It is important to point out that the examples of [106] with the Penrose diagram of Figure 8 have a non-singular Cauchy horizon $\mathscr{CH}^+$ (consistently with Theorem 5.3), a situation which is conjectured to be non-generic. Indeed, in Theorem 4.7, we recall that $\mathscr{CH}^+$ was shown to be weakly singular under decay assumptions on the event horizon that are conjectured to be generic. The main mechanism in Theorem 5.3—the "breakdown of the weak null singularity" $\mathscr{CH}^+$—is proved by a contradiction argument, using the focusing properties of the Einstein equations (notably the null constraint, so-called Raychaudhuri, equations) and taking advantage of the strength of the singularity at $\mathscr{CH}^+$ combined with the spacetime geometry near the central axis Γ. In view of the breakdown of $\mathscr{CH}^+$, there must be another (presumably spacelike, see Figure 9) component of the terminal boundary; we will discuss this aspect in Section 5.4 below.

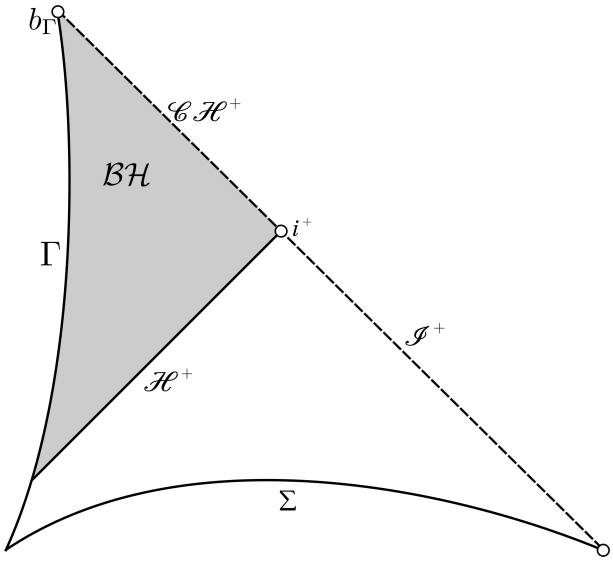


Figure 8. Impossible Penrose diagram if $\mathscr{CH}^+$ is weakly singular by Theorem 5.3.

5.4. *The spacelike portion of the terminal boundary*

Theorem 5.3 is consistent with a spacetime geometry whose Penrose diagram is the rightmost in Figure 4, i.e., with a Cauchy horizon $\mathscr{CH}_\Gamma$ emanating from a locally naked singularity b_Γ. Barring this possibility, however, Theorem 5.3 shows that $\mathscr{B} - \mathscr{CH}^+ = \mathscr{S} = \{r = 0\}$, where $\mathscr{S}$ is not entirely null and presumably spacelike [87]. The conjectured non-genericity of locally naked singularities discussed in Section 5.2 together with Theorem 5.3 then imply that a generic spherically-symmetric black hole solution of the Einstein–Maxwell-charged-scalar-field system (4.4) has the Penrose diagram of Figure 9, as was originally conjectured by Dafermos [26]. Note the presence of two co-existing singular components of the terminal boundary of different strengths, namely $\mathscr{CH}^+$, which is a weak singularity (blow-up of curvature) and $\mathscr{S} = \{r = 0\}$, which is a stronger singularity. The quantitative study of the transition region between $\mathscr{CH}^+$ and $\mathscr{S}$ in Figure 9 was carried out in the author's recent work [95], under the assumption that the scalar field oscillates weakly[19] towards the Cauchy horizon $\mathscr{CH}^+$. The main result is that $\mathscr{S} = \{r = 0\}$ is indeed spacelike[20] in the vicinity of $\mathscr{CH}^+$ and described by a Kasner-type metric with variable Kasner exponents (see [95,107] for a definition) that degenerate towards the intersection $\mathscr{CH}^+ \cap \mathscr{S} \subset \{v = \infty\}$, where v is the same advanced-time coordinate as in (4.6).

Theorem 5.4 ([95]). *Assume the spacetime in Theorem 5.3 does not possess a locally naked singularity, i.e., $\mathscr{CH}_\Gamma = \emptyset$, and denote $\mathscr{B}$ its terminal boundary. Then $\mathscr{B}$ contains a spacelike singularity $\mathscr{S} \neq \emptyset$ intersecting $\mathscr{CH}^+ \subset \{v = \infty\}$ and the metric near $\mathscr{CH}^+ \cap \mathscr{S}$ is approximated by a Kasner metric of v-dependent Kasner exponents $(1-2p(v), p(v), p(v))$ with $p(v) \approx v^{-1}$ as $v \to \infty$.*

Remark. Strictly speaking, the analysis of Theorem 5.4 is unnecessary to prove a spherically-symmetric version of Strong Cosmic Censorship for the charged collapse model, since [13] the only possible boundary components in spherical symmetry are the Cauchy horizon $\mathscr{CH}^+$ and boundary components on which $r = 0$ (excluding the conjecturally non-generic locally naked singularities $\mathscr{CH}_\Gamma$). C^2 extensions through $\mathscr{CH}^+$ are impossible by Theorem 4.7, and through

[19] Note that this assumption is always satisfied in the uncharged scalar field model (4.2), since ϕ is real-valued.

[20] In particular, there is no null ingoing boundary component $\mathscr{S}_{i^+}$ on which $r = 0$ and connected to $\mathscr{CH}^+$, see [95].

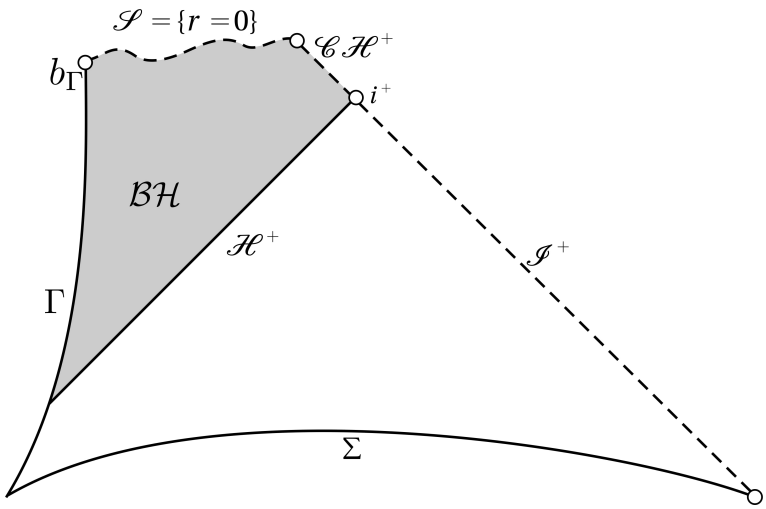


Figure 9. Penrose diagram of the one-ended black holes of Theorem 5.3, with $\mathscr{S} = \{r = 0\}$.

components on which $r = 0$, due to the blow-up of the Kretschmann scalar [13]. However, a complete resolution of Strong Cosmic Censorship *without spherical symmetry* requires a comprehensive quantitative understanding of the co-existence of weak and strong singularities depicted in Figure 9.

Theorem 5.4 provides the first examples of a null-spacelike singularity transition[21] in a black hole, together with the first construction a Kasner-like singularity with degenerating exponents (see [107] for a very general construction of Kasner-type spacetimes, which however assumes the Kasner exponents to be non-degenerating and thus does not cover the situation of Theorem 5.4). Lastly, one may speculate that the results of Theorem 5.4 can be suitably generalized[22] to non spherically-symmetric spacetimes solving the Einstein-scalar-field system (2.4), (3.7). In summary, it is conjectured that the geometry of generic (non spherically-symmetric) Einstein-scalar-field solutions is as in Figure 9. The situation for the Einstein vacuum equations, however, is more uncertain in that Kasner-type singularities are not expected to be stable, see [107,108].

6. Late-time tails in the black hole exterior and connections to black hole stability

In this section, we discuss decay results on the event horizon $\mathscr{H}^+$ of the form (4.6). We have emphasized in Section 4 that decay assumptions are essential in the black interior dynamics (recalling Theorems 4.4, 4.5, 4.6 and 4.7). In particular, upper bounds are used in stability results of the Cauchy horizon $\mathscr{CH}^+$ and lower bounds to show its (weakly) singular character. Proving Strong Cosmic Censorship starting from generic initial data in the black hole exterior, as prescribed in Conjecture 1.1, requires proving that such upper and lower bounds are satisfied on the event horizon $\mathscr{H}^+$ for generic (smooth) initial data. We discuss results in this direction below, starting from the toy problem provided by linear wave-type equations on a fixed black hole.

[21] Theorem 5.4 is, in fact, the corollary of a more general local result in [95]. In particular, the Kasner-like quantitative estimates of Theorem 5.4 also apply to the two-ended black holes with $\mathscr{S} \neq \emptyset$ as in the rightmost diagram of Figure 7.

[22] The landmark result of Fournodavlos–Rodninanski–Speck [92] indeed proves that Kasner metrics with constant positive exponents are stable to small perturbations. The Kasner exponents in Theorem 5.4 are also positive, thus the spacetime is expected to be stable against non-spherical perturbations. However, the Kasner exponents in Theorem 5.4 are non-constant and degenerate as $v \to +\infty$, and therefore the techniques of [92] do not apply.

6.1. *Linear wave equations on a stationary black hole and Price's law*

Inverse-polynomial time decay for the wave equation

$$\Box_g \phi = 0 \tag{6.1}$$

on a black hole spacetime has been conjectured since the numerics of Price [109] articulating what is now known as Price's law. Regarding mathematical works, boundedness and integrated decay for (6.1) on a Schwarzschild $g = g_S$ or (sub-extremal) Reissner–Nordström $g = g_{\mathrm{RN}}$ black hole were established in [110,111] and [27,42,111–114] on the (sub-extremal) Kerr black hole $g = g_{\mathrm{K}}$. Decay upper bounds were proven on Schwarzschild/Reissner–Nordström [115,116] and Kerr [113,117]. Precise asymptotic tails were subsequently obtained by Hintz [118] and Angelopoulos–Aretakis–Gajic [119,120]. While these tails are valid throughout the black hole exterior, we will only state them on the event horizon $\mathscr{H}^+$ in the theorem below for simplicity, using an Eddington–Finkelstein advanced-time coordinate v on $\mathscr{H}^+$ as in Section 4.

Theorem 6.1 (Price's law, [118–120]). *Let g be either a Schwarzschild $g = g_S$ or (sub-extremal) Reissner–Nordström black $g = g_{\mathrm{RN}}$ or (sub-extremal) Kerr $g = g_{\mathrm{K}}$ black hole exterior and ϕ a solution of (6.1) with smooth compactly supported initial data at $\{t = 0\}$. Then, there exists a function $I_0(\theta,\varphi)$ on $\mathbb{S}^2$, $\epsilon > 0$, such that ϕ obeys the following asymptotics as $v \to +\infty$:*

$$\phi_{|\mathscr{H}^+}(v,\theta,\varphi) = I_0(\theta,\varphi)\, v^{-3} + O(v^{-3-\epsilon}). \tag{6.2}$$

Moreover, $I_0(\theta,\varphi) \not\equiv 0$ for generic initial data.

The proof of (6.2) (and estimates for higher order derivatives) in Theorem 6.1 show that the assumptions of Theorem 4.2 are satisfied for generic smooth and localized solutions of (6.1).

While (6.1) is often considered as a linear toy model for the Einstein equations (2.4) in a vacuum, the equations of linearized gravity around the Kerr metric (3.5) give rise to the celebrated Teukolsky equations [121] which are spin-weighted analogues of (6.1) and have the following form on (sub-extremal) Kerr spacetime (3.5), where the spin s takes the value $\{\pm 2\}$.

$$\begin{aligned}\Bigg[\Box_{g_{\mathrm{K}}} &+ \frac{2s}{r^2 + a^2\cos^2(\theta)}(r-M)\partial_r + \frac{2s}{r^2+a^2\cos^2(\theta)}\left(\frac{a(r-M)}{(r-r_+)(r-r_-)} + i\frac{\cos(\theta)}{\sin^2(\theta)}\right)\partial_\varphi \\ &+ \frac{2s}{r^2+a^2\cos^2(\theta)}\left(\frac{M(r^2-a^2)}{(r-r_+)(r-r_-)} - r - \mathrm{i}a\cos(\theta)\right)\partial_t + \frac{1}{r^2+a^2\cos^2(\theta)}(s - s^2\cot^2(\theta))\Bigg]\alpha_s = 0.\end{aligned} \tag{6.3}$$

Boundedness and integrated decay for (6.3) have been established in [122–126], and late-time tail results analogous to Theorem 6.1 were obtained in [127–129]; see [130–135] for analogous results on Schwarzschild/Reissner–Nordström black holes (namely, the analogue of (6.3) with $a = 0$).

6.2. *Charged scalar field equation*

Motivated by the study of the Einstein–Maxwell-charged-scalar-field equations discussed in Sections 4 and 5, we introduce the charged scalar field equivalent of (6.1) on a fixed black hole consisting of the following nonlinear Maxwell-charged-scalar-field system of equations:

$$g^{\mu\nu}D_\mu D_\nu \phi = 0, \quad \nabla^\mu F_{\mu\nu} = \mathrm{i}q_0 \Im(\bar{\phi} D_\nu \phi), \quad D_\mu = \nabla_\mu + \mathrm{i}q_0 A_\mu, \quad F = \mathrm{d}A. \tag{6.4}$$

For spherically-symmetric solutions of (6.4) on a black hole metric g (like (3.2) or (3.3)), we define the local charge Q (a function on spacetime) and the asymptotic charge $e \in \mathbb{R}$ as follows:

$$e = \lim_{v\to+\infty} Q_{|\mathscr{H}^+}(v), \quad \text{where } F = \frac{Q}{r^2}\Omega^2 \mathrm{d}u \wedge \mathrm{d}v \text{ and } g = -\Omega^2 \mathrm{d}u\mathrm{d}v + r^2(\mathrm{d}\theta^2 + \sin^2(\theta)\,\mathrm{d}\varphi^2). \tag{6.5}$$

One should expect $e \neq 0$ generically. As it turns out, for $g = g_S$ or $g = g_{\mathrm{RN}}$ (in the sub-extremal case), $g^{\mu\nu} D_\mu D_\nu \phi = 0$ asymptotically resembles a wave equation with an inverse-square potential [136,137], where the potential is of the schematic form $V(r) \approx \mathrm{i} q_0 \mathrm{e}\, r^{-2}$. Heuristics in the physics literature [75,76] have led to the following conjectured asymptotic tail for solutions of (6.4)

$$\phi_{|\mathscr{H}^+}(v,\theta,\varphi) = C_H \cdot \mathrm{e}^{\frac{\mathrm{i} q_0 e}{r_+} v} \cdot v^{-1-\sqrt{1-4(q_0 e)^2}} + O(v^{-1-\epsilon}), \tag{6.6}$$

where $C_H \neq 0$ for generic solutions and $\epsilon > 0$. Note that the decay condition of the second part of Theorem 4.4 is satisfied by (6.6), as is the oscillation condition (4.8) required in Theorem 4.5. While the proof of (6.6) is still open for spherically-symmetric solutions of (6.4), sharp upper bounds corresponding to (6.6) have been established in the limit $|q_0 e| \ll 1$ by the author [136].

Theorem 6.2 ([136]). *Spherically-symmetric solutions of (6.4) on a sub-extremal Reissner–Nordström metric $g = g_{\mathrm{RN}}$ with smooth, localized and small initial data at $\{t = 0\}$ obey decay upper bounds of the form (4.6) for some $s(q_0 e) > 1$, and $s(q_0 e) \to 2$ as $q_0 e \to 0$.*

We also refer to [137] for the proof of late-time tails of the form (6.6) for a toy-model of (6.4).

6.3. *The Klein–Gordon equation for massive scalar fields*

We now turn to the Klein–Gordon equation (4.5) (with $m^2 > 0$), which is the massive variant of the wave equation (6.1) (noting that Theorems 4.4, 4.5 and 4.7 still apply to this case). We emphasize that the picture of Strong Cosmic Censorship for the Einstein–Klein–Gordon equations is conjectured to be qualitatively different from the vacuum case. Some solutions of (4.5) on a (sub-extremal) Kerr spacetime indeed do not decay, but instead grow exponentially in time (growing mode solution), as proven by Shlapentokh-Rothman [69] (see also [138,139] for pioneering works in the Physics literature, together with the review [140]).

Theorem 6.3 (Kerr superradiant instability, [69]). *Fix a sub-extremal Kerr (3.5) metric. For an open set of $m^2 > 0$, there exist growing mode solutions of (4.5) with smooth, localized initial data.*

In view of the instability of the Kerr metric, one would need to identify an (hypothetical) generic endstate[23] of Klein–Gordon collapse and show its inextendibility as a first step to proving Conjecture 1.1, a problem which is completely open at present. However, the Klein–Gordon equation (4.5) on a Schwarzschild (3.2) or (sub-extremal) Reissner–Nordström (3.3) black hole, in contrast to the Kerr case, decays on the event horizon $\mathscr{H}^+$ at an inverse-polynomial rate of the form (4.6) with $s = 5/6$ as proven by Pasqualotto, Shlapentokh-Rothman and the author [70,143].

Theorem 6.4 ([70,143]). *Let ϕ be a solution of the Klein–Gordon equation (4.5) with $m^2 > 0$ on a fixed Schwarzschild (3.2) or (sub-extremal) Reissner–Nordström (3.3) black hole with smooth and compactly supported initial data at $\{t = 0\}$. Then for every $\epsilon > 0$, there exists $D > 0$ such that:*

$$|\phi_{|\mathscr{H}^+}|(v,\theta,\varphi) \leq D \cdot v^{-\frac{5}{6}+\epsilon}.$$

Moreover, there exists $C \in \mathbb{R}$ such that, if ϕ has spherically-symmetric initial data, for every $\delta > 0$

$$\left|\phi_{|\mathscr{H}^+}(v,\theta,\varphi) - F(v)\, v^{-\frac{5}{6}}\right| \leq D \cdot v^{-1+\delta},$$

$$F(v) = C \textstyle\sum_{q=1}^{+\infty} \gamma^{q-1} q^{\frac{1}{3}} \cos(\tilde{\Psi}_q(v)), \quad \tilde{\Psi}_q(v) = m v - \frac{3}{2}[2\pi M]^{\frac{2}{3}} m q^{\frac{2}{3}} v^{\frac{1}{3}} + \phi_-(M,e,m^2) + O(v^{-\frac{1}{3}}), \tag{6.7}$$

where $|\gamma|(M,e,m^2) < 1$ and $\phi_-(M,e,m^2) \in \mathbb{R}$. Moreover, $C \neq 0$ for generic initial data.

[23] It has been debated [138,139,141,142] whether the Schwarzschild black hole may be a possible endstate. Note also that the non-trivial stationary black holes solutions of the Einstein–Klein–Gordon equations (so-called "hairy black holes") constructed by Chodosh–Shlapentokh-Rothman [71,72] are also hypothetical endstate candidates. Lastly, it is also possible that *some* sub-extremal Kerr black hole are immune to the superradiant instability, and thus could be stable [69].

In the spherically symmetric case, $\phi_{|\mathscr{H}^+}$ satisfies decay upper and lower bounds analogous to the ones required in Theorem 4.4 and Theorem 4.7; moreover, the leading order term $F(v)v^{-5/6}$ satisfies the oscillation condition (4.8) (with $\omega_{\text{res}} = 0$, since $q_0 = 0$ in Theorem 6.4) corresponding to the one required in Theorem 4.5. In view of Theorem 6.4, we conjecture that (6.7) also holds for spherically-symmetric solutions for the Einstein–Klein–Gordon model with regular and localized initial data; proving such a statement is indeed key to a complete resolution of the spherically-symmetric version of Strong Cosmic Censorship for the (massive) Einstein–Klein–Gordon model.

6.4. *Wave equations on dynamical black holes and black hole stability*

An important pre-requisite to Strong Cosmic Censorship in the form of Conjecture 1.1 is the proof of asymptotic stability of the black hole exterior region. In the simpler setting of spherical symmetry, this was achieved by Dafermos–Rodnianski [74] for the Einstein–Maxwell-scalar-field system (4.2), who also proved upper bounds (Price's law) analogous to (6.2) on the event horizon that are sufficient to satisfy the assumptions of Theorem 4.4. Luk–Oh [84] subsequently proved that (L^2-averaged) inverse-polynomial lower bounds hold on the event horizon, thus showing that the assumptions of Theorem 4.7 are satisfied for spherically-symmetric solutions of (4.2) with generic, smooth and localized initial data at $\{t = 0\}$. Recently Luk–Oh [144] obtained asymptotic point-wise tails (hence point-wise lower bounds). We summarize these results below.

Theorem 6.5 ([74,84,144]). *Given any asymptotically flat, spherically symmetric, admissible data on $\mathbb{R} \times \mathbb{S}^2$, the MGHD $(\mathscr{M}, g, \phi)$ for (4.2) converges to a Reissner–Nordström black hole at the following rate: there exists $D \in \mathbb{R}$, $\epsilon > 0$, $C \in \mathbb{R}$ such that*

$$\left|\phi_{|\mathscr{H}^+}(v,\theta,\varphi) - D \cdot v^{-3}\right| \leq C \cdot v^{-3-\epsilon}. \tag{6.8}$$

Moreover, $D \neq 0$ for generic initial data.

Remark. Note that no smallness assumption is imposed on the initial data in Theorem 6.5. However, the admissibility condition requires the presence of a trapped surface at $\{t = 0\}$, therefore the dynamical solution cannot disperse to the Minkowski spacetime.

Returning to the vacuum Einstein equations without symmetry, the proof of stability of the Kerr black hole exterior (3.5) was obtained in the case $|a| \ll M$ in the work of Giorgi, Klainerman, Shen and Szeftel [77,78,145–148]; see the closely related work of Dafermos–Holzegel–Rodnianski–Taylor [149] on the stability of Schwarzschild, as well as [112,122–124,135,150–152] for related linear and nonlinear black hole stability results. The decay upper bounds obtained on the event horizon $\mathscr{H}^+$ of the dynamical Kerr perturbations satisfy the assumptions of Theorem 4.6: therefore, small perturbations of the Kerr exterior (3.5) (for $|a| \ll M$) admit a C^0-extendible Cauchy horizon as depicted in Figure 6. Linearized results [122,123] on solutions of (6.3) indicate that the assumptions of Theorem 4.6 are also satisfied for small perturbations of the Kerr exterior (3.5) in the full sub-extremal range $|a| < M$. We emphasize, however, that a full resolution of Conjecture 1.1 requires to understand the generic endstate of gravitational collapse, which may involve several black holes. In that respect, we mention the celebrated *final state conjecture* [1], postulating that a generic solution of (2.4) consists of finitely many Kerr black holes moving away from each others. Understanding spacetimes with multiple black-hole exterior regions is thus a necessary preliminary step prior to a full proof of Strong Cosmic Censorship.

We have already emphasized that decay lower bounds on a black hole's event horizon are necessary to prove the (weakly) singular character of Cauchy horizon $\mathscr{CH}^+$: proving such lower bounds naturally requires to understand the late-time tail of wave equations on a dynamical black hole, an endeavor going beyond the problems of Section 6.1. For (6.1), it is known that the Price's law asymptotics of Theorem 6.1 also hold on a class of dynamical black holes [117,144].

However, it turns out that for other wave equations, the late-time tails are in general *different on dynamical black holes* (at least in odd space-dimensions) as recently established by Luk–Oh [144]. While their result provides late-time asymptotics for a general class of wave equations in odd space dimensions, for concreteness, we focus on the following example of (6.1) on a spherically-symmetric dynamical black hole $\tilde{g}$ converging (sufficiently fast) to the Schwarzschild exterior g_S (in $(3+1)$ spacetime dimensions). Defining $\Pi_{\geq \ell}$ as the projection on spherical harmonics of order ℓ and higher and assuming compactly supported initial data, the result of Luk–Oh [144] provides the following dichotomy, under the assumption that the metric $\tilde{g}$ is non-stationary and $\ell \geq 1$:

- If $\Box_{\tilde{g}}\phi = 0$, then $\Pi_{\geq \ell}\phi_{|\mathscr{H}^+}(v,\theta,\varphi) \sim C_\ell(\theta,\varphi)\, v^{-2-2\ell}$, and $C_\ell \not\equiv 0$ for generic initial data.
- If $\Box_{g_S}\phi = 0$, then $\Pi_{\geq \ell}\phi_{|\mathscr{H}^+}(v,\theta,\varphi) \sim C_\ell(\theta,\varphi)\, v^{-3-2\ell}$, and $C_\ell \not\equiv 0$ for generic initial data.

The asymptotics of higher order spherical harmonics have no direct relevance to Strong Cosmic Censorship. However, there is a strong resemblance between by $\Pi_{\geq 2}\phi$, where ϕ solves (6.1) and the Teukolsky equation (6.3) for $s = \pm 2$. Motivated by this analogy, Luk–Oh conjectured [144] that generic black hole solutions of the vacuum Einstein equations approaching a sub-extremal Kerr metric decay at the rate v^{-6} on the event horizon $\mathscr{H}^+$, as opposed to the v^{-7} rate proven for solutions of (6.3) on the exactly stationary Kerr exterior [129]. Proving such asymptotic tails on a dynamical black hole will be a crucial step towards a proof of Conjecture 1.1.

7. Strong Cosmic Censorship conjecture beyond the context of gravitational collapse

We have focused so far on aspects of Strong Cosmic Censorship in generic $(3+1)$-dimensional gravitational collapse (except in Sections 4.6 and 4.7), the context in which it was originally formulated by Penrose [4]. In this section, we discuss Strong Cosmic Censorship in other contexts, together with related questions that are in principle excluded by the genericity condition in Conjecture 1.1, but still relate to the dynamics of black holes and their interior singularities.

7.1. *Exceptional dynamical black holes arising from finite co-dimension initial data*

We start remarking that the one-parameter Schwarzschild family (3.2) is a co-dimension three (respectively one) sub-family of the Kerr family (3.5) (respectively the Reissner–Nordström family (3.3)) corresponding to $a = 0$ (respectively $e = 0$). Similarly, the extremal Reissner–Nordström ($|e| = M$) and Kerr ($|a| = M$) solutions have co-dimension one within their respective families. In view of this, it is reasonable to conjecture that black holes converging to an extremal Reissner–Nordström/Kerr solution (asymptotically extremal black holes) or Schwarzschild (asymptotically Schwarzschildean black holes) have non-generic initial data within the moduli space of regular solutions, which is why they should not play a major role in a proof of Conjecture 1.1.

Asymptotically extremal black holes. The recent constructions of Kehle–Unger [106,153] suggest that asymptotically extremal black holes indeed arise for a finite codimensional subset of asymptotically flat initial data, at least within spherically symmetric models. Spherically-symmetric asymptotically extremal black hole solutions of (4.4) (with a small charge) were studied by Gajic–Luk [154] (see also previous linear results by Gajic [155,156]) showing that under a decay condition like (4.6) on the event horizon, the spacetime admits a non-trivial Cauchy horizon $\mathscr{CH}^+$ (like in Theorem 4.4) and the metric is $C^0 \cap H^1_{\text{loc}}$-extendible across the Cauchy horizon (in contrast to Theorem 4.7). This shows that one can extend the metric as a weak solution to the Einstein–Maxwell-charged-scalar-field system of equations (recall the discussion in Statement (II) of Section 2.3). An analogous statement for asymptotically extremal solutions of the Einstein vacuum equations (without any symmetry assumptions) remains open.

Asymptotically Schwarzschildean black holes. Within spherical symmetry, black hole solutions of (3.7) with localized initial data are all asymptotically Schwarzschildean (this follows from [74]) and the MGHD features a spacelike singularity $\mathscr{S} = \{r = 0\}$, as discussed in Section 5.2. For spherically-symmetric models with charge, in which asymptotically Schwarzschildean black holes are non-generic, it is shown in the author's [157] that asymptotically Schwarzschildean spherically-symmetric solutions of (4.4) have no Cauchy horizon ($\mathscr{CH}^+ = \emptyset$) and the MGHD terminal boundary is foliated by spheres of area-radius $r = 0$. For the Einstein vacuum equations, the result of Dafermos–Holzegel–Rodnianski–Taylor [149] shows that asymptotically Schwarzschildean black holes arise from a co-dimension three subset of initial data close to Schwarzschild. For these asymptotically Schwarzschildean black holes, the spacetime supremum of Kretschmann curvature blows up [22,149] in the black hole interior, which suggests that the MGHD is future-inextendible. However, the causal nature (null or spacelike) of the singularity, and quantitative estimates in its vicinity, however, remain open, even within spherical symmetry.

7.2. *Spatially-compact spacetimes and cosmological analogues of Cosmic Censorship*

Hawking singularity theorem and timelike incompleteness. Spacetimes admitting a *compact Cauchy surface* are particularly relevant to Cosmology as models of a finite universe and are subjected to the celebrated Hawking singularity theorem [158] (compare to Theorem 2.5 which is, in contrast, more suitable to the context of gravitational collapse).

Theorem 7.1 ([158]). *If $(\mathscr{M}, g)$ admits a compact Cauchy hypersurface with positive mean-curvature and satisfies the energy condition (2.6), then it is future-timelike geodesically incomplete.*

Similarly to Theorem 2.5, however, Theorem 7.1 does not provide any information on the cause of geodesic incompleteness or the singular nature of the MGHD terminal boundary.

The rigidity of compact Cauchy horizons. As in the gravitational collapse case, Cauchy horizons[24] can be a cause for MGHD extendibility, and thus constitute the main obstruction to Strong Cosmic Censorship. A possible proof strategy of Strong Cosmic Censorship for spatially-compact spacetimes is to show that for solutions to the Einstein vacuum equations, any *compact Cauchy horizon is a Killing horizon*, as conjectured by Isenberg–Moncrief [159]. Such ridigity results show that Cauchy horizons only exist in spacetimes with an extra symmetry and thus are non-generic. See [160,161] for a partial resolution of the Isenberg–Moncrief problem.

Strong Cosmic Censorship in vacuum under Gowdy-symmetry. Strong Cosmic Censorship was studied for a class of spacetimes with a compact spatial topology $\mathbb{T}^3$, $\mathbb{S}^3$ or $\mathbb{S}^2 \times \mathbb{S}^1$ admitting a two-dimensional isometry group $\mathbb{U}(1) \times \mathbb{U}(1)$ with spacelike orbits, under an additional two-surface orthogonality condition (Gowdy-symmetry, see [162–164]). The Gowdy polarized subcase is defined as the instance in which the two independent Killing vector fields generated by the isometry group are orthogonal. In view of Theorem 7.1, one may expect these spacetimes to feature a spacelike singularity at $\{t = 0\}$. A version of Strong Cosmic Censorship within polarized-Gowdy solutions of the Einstein vacuum equations was proven by Chrusciel–Isenberg–Moncrief [19] for any $\mathbb{T}^3$, $\mathbb{S}^3$ or $\mathbb{S}^2 \times \mathbb{S}^1$ topology. Strong Cosmic Censorship within the more general unpolarized-Gowdy class was subsequently proved by Ringström [165,166] in the $\mathbb{T}^3$ topology. Both unpolarized and polarized results proceed proving the blow-up of the Kretschmann scalar along incomplete geodesics, employing asymptotic expansions in the vicinity of the singularity at $\{t = 0\}$. The unpolarized case is significantly more involved on a technical level due

[24] As an example, the Taub–NUT spacetime has both a compact Cauchy surface and an extendible Cauchy horizon [6].

to the presence of so-called *spikes* [164,165] which occur when the limit of some renormalized metric components (more precisely, their time-derivative) as $t \to 0$ is a discontinuous function on $\{t = 0\}$. A key element of the proof in [165,166] is to construct a subset of regular initial data $\mathscr{G}$ for which only finitely-many spikes occur in evolution and such that $\mathscr{G}$ is both dense in a smooth topology and open in a (possibly weaker) regular topology (this is the definition of $\mathscr{G}$ being a "generic set").

Strong Cosmic Censorship with collisionless matter under surface symmetry. Dafermos–Rendall [167–169] and Smulevici [170] considered solutions of the Einstein–Vlasov equations for surface-symmetric spacetimes with topology $\mathbb{S}^1 \times \Sigma$, where Σ is a 2-dimensional compact surface of constant curvature $k \in \{-1, 0, 1\}$ (note that the Gowdy symmetry class with $\mathbb{S}^2 \times \mathbb{S}^1$ topology is a subclass of the $k = 1$ case). They established the C^2 inextendibility of the MGHD (both for vanishing $\Lambda = 0$ and positive $\Lambda > 0$ cosmological constant), thus proving a version of Strong Cosmic Censorship for collisionless matter within the surface-symmetry class. It is remarkable that, contrary to Gowdy-symmetric vacuum spacetimes, precise asymptotics near the singularity are not essential in the case of collisionless matter considered in [167–170] and inextendibility instead follows from softer geometric arguments, in particular the extension principle of [168].

Strong Cosmic Censorship in the vicinity of Kasner spacetimes. Recall the Kasner metric (4.10) from Section 4.7 is an anisotropic spatially-homogeneous spacetime with compact spatial topology $\mathbb{T}^3$ and a spacelike singularity at $\{t = 0\}$. Its perturbations for the Einstein vacuum equations may have very intricate dynamics (BKL heuristics [93,94,171]). However, in the presence of a scalar field (i.e., for solutions of (2.4), (3.7)) or within $\mathbb{U}(1)$-polarized solutions of the Einstein vacuum equations, there exist Kasner metrics (4.10) with positive exponents (p_1, p_2, p_3). The work of Fournodavlos–Rodnianski–Speck [92] proves the stability of such solutions, and one can thus infer a version of Strong Cosmic Censorship in a vicinity of these Kasner spacetimes for the Einstein-scalar-field equations (or the vacuum Einstein equations in $\mathbb{U}(1)$-polarized symmetry). The results of [92] were recently generalized by Groeninger–Peterson–Ringström [172] proving the stability of a larger class of so-called "quiescent singularities" that includes Kasner metrics (4.10) with positive exponents p_i. Thus, an analogous statement of Strong Cosmic Censorship can be obtained in the vicinity of the larger class of initial data considered in [172].

7.3. *Strong Cosmic Censorship in higher-dimensional gravity*

Higher-dimensional black hole stability. There are higher-dimensional analogues of the Schwarzschild (3.2) and Kerr (3.5) black holes, so-called Myers–Perry black holes [173]. The techniques used for black hole exterior stability in $(3+1)$-dimensions discussed in Section 6 do not apply to the higher-dimensional case; see, however, [174–178] for interesting recent progress.

Extra stationary black holes: black rings. The Schwarzschild and Myers–Perry solutions are not the only stationary asymptotically-flat higher-dimensional black holes in view of the celebrated black ring solutions [179], which possess an event horizon with $\mathbb{S}^2 \times \mathbb{S}^1$ topology. However, the black ring is subjected to the celebrated Gregory–Laflamme instability [180] taking the form of exponentially growing modes for (6.1), see [181] for a proof of this instability in $(4+1)$ dimensions. Additionally, stable trapping takes place for (6.1) if g is a black ring, leading to a slow logarithmic decay-in-time [182], a potential source of further (nonlinear) instabilities.

Open problem. To summarize, even the conjectured endstate of higher-dimensional gravitational collapse appears to be more intricate than its $(3+1)$-dimensional counterpart; thus, an hypothetical higher-dimensional analogue of Conjecture 1.1 remains completely open at present.

7.4. *Non-vanishing cosmological constant*

Positive cosmological constant. The Schwarzschild-de-Sitter and Kerr-de-Sitter (respectively Reissner–Nordström-de-Sitter) spacetimes are black hole solutions of the Einstein vacuum (respectively Einstein–Maxwell) equations with a positive cosmological constant $\Lambda > 0$. Their Cauchy surface (for a region inside the cosmological horizon, thus excluding the cosmological region, see e.g. [42]) is compact with the topology of $\mathbb{S}^2 \times \mathbb{S}^1$, thus the de-Sitter black holes are not asymptotically flat (their interior region, however, is similar to their $\Lambda = 0$ counterparts).

Consequently, solutions of the wave equation (6.1) do not decay at an inverse-polynomial rate (in the style of (4.6)) on the event horizon, but instead at an exponential rate [183–186], see also [187–189] for such results on nonlinear wave equations. At the level of the Einstein-vacuum equations, the stability of the Kerr-de-Sitter black holes (with $|a| \ll M$) at an exponential rate was proven in the celebrated work of Hintz–Vasy [190], reproved in [191]. We note that Theorem 4.6 only requires decay upper bounds on the event horizon (compatible with an exponential rate) and thus applies mutatis mutandis to small perturbations of Kerr-de-Sitter black holes with $|a| \ll M$, which therefore possess a C^0-extendible Cauchy horizon.

However, due to the fast exponential decay in the $\Lambda > 0$ case, the singular character of the Cauchy horizon, and thus the validity of the H^1 formulation of Strong Cosmic Censorship appears in doubt (see [192] and the appendix of [21]). Heuristics and numerical works on the wave equation (6.1) indeed indicate that solutions with smooth and localized initial data are actually H^1_{loc}-bounded at the Cauchy horizon of a Reissner–Nordström-de-Sitter hole [193] or Kerr–Newman-de-Sitter [194] for sub-extremal parameters sufficiently close to extremality (in contrast to the combined result of Theorem 6.1 and Theorem 4.2 in the $\Lambda = 0$ case); however, on a Kerr-de-Sitter[25] black hole, solutions to (6.1) seem to blow-up in H^1 norm at the Cauchy horizon [195].

At the nonlinear level, spherically-symmetric Einstein–Maxwell-(charged)-scalar-field solutions of (4.4) approaching a (sub-extremal) Reissner–Nordström-de-Sitter black hole were studied in [196–198]. Under the assumption of exponential decay of the scalar field on the event horizon, the stability of the Cauchy horizon was obtained analogously to Theorem 4.4. For a sufficiently rapidly-decaying scalar field on the event horizon, however, the Cauchy horizon is not weakly singular and the Hawking mass remains bounded in this case (in complete contrast to Theorem 4.7 obtained in the $\Lambda = 0$ case). Whether or not generic spherically-symmetric solutions decay sufficiently fast on the event horizon for this scenario to materialize remains open, however.

Negative cosmological constant. On the Schwarzschild/Kerr-anti-de-Sitter black hole (solutions of the vacuum Einstein equations with $\Lambda < 0$ analogous to (3.2) and (3.5)), solutions of the wave equation (6.1) decay at a slow inverse-logarithmic rate [199,200]. On a Reissner–Nordström-anti-de-Sitter black hole, a recent result of Zheng [201] shows the existence of non-trivial stationary solutions of (6.1) (see also previous heuristics works [202,203] in the context of AdS-CFT) and exponentially-growing modes (the former also being associated to asymptotically-AdS hairy black holes [90]). Excluding these, Kehle proved that solutions to (6.1) are (point-wise) bounded at the Reissner–Nordström-anti-de-Sitter Cauchy horizon [204], although generically their H^1 norm blows-up [205]. On a Kerr-anti-de-Sitter black hole, in contrast, Kehle surprisingly showed [206] that solutions of (6.1) blow up point-wise at the Cauchy horizon, for black hole parameters $(M, |a|, \Lambda)$ belonging to a Baire-generic subset (of zero Lebesgue measure) of $\mathbb{R}^3_+$. The exact repercussions of these instabilities on the Einstein-vacuum dynamics, however, are unclear in view of the conjectured nonlinear instability of the Kerr-anti-de-Sitter exterior spacetime [207].

[25] The numerics of [193–195] thus surprisingly suggest that Strong Cosmic Censorship (at the H^1 level, i.e., Statement (II) of Section 2.3) for $\Lambda > 0$ may be violated in the presence of charge (Maxwell field), but respected in its absence.

8. Summary and future directions

This review explores the burgeoning field of Strong Cosmic Censorship research, which has developed its own momentum within the mathematics landscape of General Relativity, while remaining faithful to Penrose's original spirit. The understanding of its key features were indeed made possible thanks to advancements in the theory of quasilinear waves, started in the 1990s, which drove immense progress in the study of spacetime dynamics in General Relativity.

We highlight a few of these essential realizations which are all connected to breakthroughs in the understanding of black hole dynamics, towards a resolution of Problem 1.2:

- The Kerr Cauchy horizon is stable against dynamical perturbations, thus the singularity inside a generic black hole arising in gravitational collapse is not entirely-spacelike.
- The blue-shift instability, a linear mechanism on a stationary black hole with a smooth Cauchy horizon, does not destroy the Cauchy horizon when backreaction is included, but instead makes it weakly singular (proven only in spherical symmetry at present).
- A black hole formed in gravitational collapse features a spacelike singularity $\mathscr{S}$ that differs significantly from the singularity inside Schwarzschild/Oppenheimer–Snyder spacetimes and whose endpoint intersects the null Cauchy horizon $\mathscr{CH}^+$ (proven only in the charged gravitational collapse of a spherically-symmetric scalar field at present).

We also emphasize the remarkable developments in the understanding of spacelike singularities dynamics. While the mathematical confirmation of the mixmaster/BKL heuristics for the Einstein vacuum equations remains an open problem, the important progress positions us well to tackle the challenges spacelike singularities pose within the resolution of Strong Cosmic Censorship.

Moreover, and despite the independence of the Weak and Strong Cosmic Censorship conjectures highlighted in Section 1.5, it is clear that a complete resolution of Conjecture 1.1 will require comprehensive knowledge of the dynamics of (locally) naked singularities in gravitational collapse, most notably their non-generic character: in other words, it is likely that a proof of Weak Cosmic Censorship will provide essential insights to the proof of Conjecture 1.1. Even a complete understanding of all the above phenomena without any symmetry assumption, however, would still not be sufficient to prove Conjecture 1.1, because it is tied to the comprehension of *every generic endstate of gravitational collapse*, and therefore the final state conjecture needs to be resolved first. Despite this obstacle, progress on the mathematical program and partial steps towards Conjecture 1.1 outlined in this review would have significant ramifications for understanding black hole dynamics and the field of General Relativity as a whole.

Among future directions, the study of multiple-black-holes dynamics, and whether they respect Strong Cosmic Censorship will be essential. Beyond the Einstein equations in vacuum, incorporating realistic astrophysical fluids, though presenting its own mathematical challenges, holds immense potential for a more comprehensive physical understanding. As a steppingstone, studying a massive (Klein–Gordon) scalar field will already reveal a drastically different picture of Strong Cosmic Censorship compared to the simpler (electro)-vacuum case.

Strong Cosmic Censorship research stands as a vibrant frontier, pushing the boundaries of both general relativity and mathematical analysis. Significant progress has been made, yet captivating challenges remain. As we delve deeper into this captivating domain, we stand to unlock a wealth of knowledge about the nature of gravity and the enigmatic realm of black holes.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

The author gratefully acknowledges the support of NSF Grant DMS-2247376.

Acknowledgments

The author acknowledges the support of NSF Grant DMS-2247376. He also thanks Haydée Pacheco for the figures in this article, and Jonathan Luk for useful comments on the manuscript.

References

[1] R. Penrose, "Some unsolved problems in classical general relativity", in *Seminar on Differential Geometry*, Annals of Mathematics Studies, Princeton University Press: Princeton, NJ, 1982, pp. 631–668.

[2] P. S. Laplace, *A Philosophical Essay on Probabilities*, French edition, Dover Publications, Inc: New York, 1995.

[3] J. Earman, *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*, The Clarendon Press, Oxford University Press: New York, 1995.

[4] R. Penrose, "Gravitational collapse: the role of general relativity", *Nuovo Cimento* **1** (1969), pp. 252–276.

[5] R. Penrose, "Singularities and time asymmetry", in *General Relativity: An Einstein Centenary Survey*, Cambridge University Press, 1980, pp. 581–638.

[6] S. W. Hawking and G. F. Ellis, *The Large Scale Structure of Space-Time*, Cambridge University Press: London-New York, 1973.

[7] B. O'Neill, *Semi-Riemannian Geometry*, Pure and Applied Mathematics, Academic Press, Inc., [Harcourt Brace Jovanovich]: New York, 1983.

[8] R. Penrose, "The question of cosmic censorship", in *Black Holes and Relativistic Stars (Chicago, IL, 1996)*, University Chicago Press: Chicago, IL, 1998, pp. 103–122.

[9] R. Penrose, "Gravitational collapse", in *Gravitational Radiation and Gravitational Collapse* (C. DeWitt-Morette, ed.), Springer, 1974, pp. 82–91.

[10] R. Geroch and G. Horowitz, "The structure of spacetime", in *General Relativity: An Einstein Centenary Survey*, Cambridge University Press, 1980, pp. 212–293.

[11] D. Christodoulou, "On the global initial value problem and the issue of singularities", *Class. Quantum Gravity* **16** (1999), no. 12A, A23–A35.

[12] K. Landsman, "Penrose's 1965 singularity theorem: from geodesic incompleteness to cosmic censorship", *Gen. Relativ. Gravit.* **54** (2022), no. 10, article no. 115.

[13] J. Kommemi, "The global structure of spherically symmetric charged scalar field spacetimes", *Commun. Math. Phys.* **323** (2013), no. 1, pp. 35–106.

[14] M. Dafermos, "Spherically symmetric spacetimes with a trapped surface", *Class. Quantum Gravity* **22** (2005), no. 11, pp. 2221–2232.

[15] R. Penrose, "Structure of space-time", in *Battelle Rencontres*, 1968, pp. 121–235.

[16] M. Simpson and R. Penrose, "Internal instability in a Reissner–Nordström black hole", *Int. J. Theor. Phys.* **7** (1973), pp. 183–197.

[17] Y. Choquet-Bruhat and R. Geroch, "Global aspects of the Cauchy problem in general relativity", *Commun. Math. Phys.* **14** (1969), pp. 329–335.

[18] V. Moncrief, "Global properties of Gowdy spacetimes with $T^3 \times \mathbf{R}$ topology", *Ann. Phys.* **132** (1981), no. 1, pp. 87–107.

[19] P. T. Chruściel, J. Isenberg and V. Moncrief, "Strong cosmic censorship in polarised Gowdy spacetimes", *Class. Quantum Gravity* **7** (1990), no. 10, pp. 1671–1680.

[20] H. Ringström, *The Cauchy Problem in General Relativity*, ESI Lectures in Mathematics and Physics, European Mathematical Society (EMS): Zürich, 2009.

[21] M. Dafermos, "Black holes without spacelike singularities", *Commun. Math. Phys.* **332** (2014), no. 2, pp. 729–757.

[22] M. Dafermos and J. Luk, *The interior of dynamical vacuum black holes I: The C^0-stability of the Kerr Cauchy horizon*, preprint, 2017, 1710.01722. to appear in *Ann. Math.*

[23] J. Luk, "Singularities in general relativity", in *ICM—International Congress of Mathematicians. Vol. 1. Sections 9–11*, EMS Press: Berlin, 2023, pp. 4120–4141. ISBN 978-3-98547-063-1; 978-3-98547-563-6; 978-3-98547-058-7.

[24] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society (EMS): Zürich, 2009.

[25] M. Dafermos, "The interior of charged black holes and the problem of uniqueness in general relativity", *Commun. Pure Appl. Math.* **58** (2005), no. 4, pp. 445–504.

[26] M. Dafermos, "Price's law, mass inflation, and strong cosmic censorship", in *7th Hungarian Relativity Workshop (RW 2003)*, 2004, pp. 79–90.

[27] M. Dafermos, I. Rodnianski and Y. Shlapentokh-Rothman, "Decay for solutions of the wave equation on Kerr exterior spacetimes III: The full subextremal case $|a| < M$", *Ann. Math. (2)* **183** (2016), no. 3, pp. 787–913.

[28] C. Kehle and M. Van de Moortel, "Strong Cosmic Censorship in the presence of matter: the decisive effect of horizon oscillations on the black hole interior geometry", *Anal. PDE* **17** (2024), no. 5, pp. 1501–1592.

[29] J. Earman, *A Primer on Determinism*, The Western Ontario Series in Philosophy of Science (WONS, volume 32), Springer, 1986.

[30] M. Dafermos, "Penrose's incompleteness theorem", *Lond. Math. Soc. Newsl.* (2021), no. 493, pp. 27–34.

[31] R. M. Wald, *General Relativity*, University of Chicago Press: Chicago, IL, 1984.

[32] Y. Choquet-Bruhat, *General Relativity and the Einstein Equations*, Oxford Mathematical Monographs, Oxford University Press: Oxford, 2009.

[33] J. Sbierski, "On the existence of a maximal Cauchy development for the Einstein equations: a dezornification", *Ann. Henri Poincaré* **17** (2016), no. 2, pp. 301–329.

[34] S. Klainerman, I. Rodnianski and J. Szeftel, "The bounded L^2 curvature conjecture", *Invent. Math.* **202** (2015), no. 1, pp. 91–216.

[35] K. A. Khan and R. Penrose, "Scattering of two impulsive gravitational plane waves", *Nature* **229** (1971), pp. 185–186.

[36] J. Luk and I. Rodnianski, "Local propagation of impulsive gravitational waves", *Commun. Pure Appl. Math.* **68** (2015), no. 4, pp. 511–624.

[37] J. Luk and I. Rodnianski, "Nonlinear interaction of impulsive gravitational waves for the vacuum Einstein equations", *Camb. J. Math.* **5** (2017), no. 4, pp. 435–570.

[38] J. Luk and M. Van de Moortel, *Nonlinear interaction of three impulsive gravitational waves I: main result and the geometric estimates*, preprint, 2021, 2301.08353.

[39] J. Luk and M. Van de Moortel, "Nonlinear interaction of three impulsive gravitational waves II: the wave estimates", *Ann. PDE* **9** (2023), no. 1, article no. 10.

[40] G. J. Galloway, E. Ling and J. Sbierski, "Timelike completeness as an obstruction to C^0-extensions", *Commun. Math. Phys.* **359** (2018), no. 3, pp. 937–949.

[41] R. Penrose, "Gravitational collapse and space-time singularities", *Phys. Rev. Lett.* **14** (1965), pp. 57–59.

[42] M. Dafermos and I. Rodnianski, "Lectures on black holes and linear waves", in *Evolution Equations*, Clay Mathematics Proceedings, American Mathematical Society: Providence, RI, 2013, pp. 97–205.

[43] *Roger Penrose – Facts – 2020*, 2024. Online at https://www.nobelprize.org/prizes/physics/2020/penrose/facts/. NobelPrize.org. Nobel Prize Outreach AB 2024.

[44] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series, Princeton University Press: Princeton, NJ, 1993.

[45] H. Lindblad and I. Rodnianski, "The global stability of Minkowski space-time in harmonic gauge", *Ann. Math. (2)* **171** (2010), no. 3, pp. 1401–1477.

[46] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, W. H. Freeman and Co.: San Francisco, CA, 1973.

[47] J. Sbierski, "The C_0-inextendibility of the Schwarzschild spacetime and the spacelike diameter in Lorentzian geometry", *J. Differ. Geom.* **108** (2018), no. 2, pp. 319–378.

[48] R. P. Kerr, "Gravitational field of a spinning mass as an example of algebraically special metrics", *Phys. Rev. Lett.* **11** (1963), pp. 237–238.

[49] J. R. Oppenheimer and H. Snyder, "On continued gravitational contraction", *Phys. Rev. (2)* **56** (1939), no. 5, pp. 455–459.

[50] D. Christodoulou, "Violation of cosmic censorship in the gravitational collapse of a dust cloud", *Commun. Math. Phys.* **93** (1984), no. 2, pp. 171–195.

[51] D. Christodoulou, "Examples of naked singularity formation in the gravitational collapse of a scalar field", *Ann. Math. (2)* **140** (1994), no. 3, pp. 607–653.

[52] D. Christodoulou, "Bounded variation solutions of the spherically symmetric Einstein-scalar field equations", *Commun. Pure Appl. Math.* **46** (1993), no. 8, pp. 1131–1220.

[53] D. Christodoulou, "The instability of naked singularities in the gravitational collapse of a scalar field", *Ann. Math. (2)* **149** (1999), no. 1, pp. 183–217.

[54] I. Rodnianski and Y. Shlapentokh-Rothman, "Naked singularities for the Einstein vacuum equations: the exterior solution", *Ann. Math. (2)* **198** (2023), no. 1, pp. 231–391.

[55] M. W. Choptuik, "Universality and scaling in gravitational collapse of a massless scalar field", *Phys. Rev. Lett.* **70** (1993), pp. 9–12.

[56] C. Gundlach, "Understanding critical collapse of a scalar field", *Phys. Rev. D* **55** (1997), pp. 695–713.

[57] S. Chandrasekhar and J. B. Hartle, "On crossing the Cauchy horizon of a Reissner–Nordström black-hole", *Proc. R. Soc. Lond. A* **384** (1982), no. 1787, pp. 301–315.

[58] M. Dafermos and Y. Shlapentokh-Rothman, “Time-translation invariance of scattering maps and blue-shift instabilities on Kerr black hole spacetimes”, *Commun. Math. Phys.* **350** (2017), no. 3, pp. 985–1016.

[59] J. M. McNamara, “Instability of black hole inner horizons”, *Proc. R. Soc. Lond. Ser. A* **358** (1978), no. 1695, pp. 499–517.

[60] J. Sbierski, “Characterisation of the energy of Gaussian beams on Lorentzian manifolds: with applications to black hole spacetimes”, *Anal. PDE* **8** (2015), no. 6, pp. 1379–1420.

[61] J. Luk and S.-J. Oh, “Proof of linear instability of the Reissner–Nordström Cauchy horizon under scalar perturbations”, *Duke Math. J.* **166** (2017), no. 3, pp. 437–493.

[62] J. Luk and J. Sbierski, “Instability results for the wave equation in the interior of Kerr black holes”, *J. Funct. Anal.* **271** (2016), no. 7, pp. 1948–1995.

[63] S. Ma and L. Zhang, “Precise late-time asymptotics of scalar field in the interior of a subextreme Kerr black hole and its application in strong cosmic censorship conjecture”, *Trans. Amer. Math. Soc.* **376** (2023), no. 11, pp. 7815–7856.

[64] J. Luk, S.-J. Oh and Y. Shlapentokh-Rothman, “A scattering theory approach to Cauchy horizon instability and applications to mass inflation”, *Ann. Henri Poincaré* **24** (2023), no. 2, pp. 363–411.

[65] J. Sbierski, “Instability of the Kerr Cauchy horizon under linearised gravitational perturbations”, *Ann. PDE* **9** (2023), no. 1, article no. 7.

[66] D. Christodoulou, “The formation of black holes and singularities in spherically symmetric gravitational collapse”, *Commun. Pure Appl. Math.* **44** (1991), no. 3, pp. 339–373.

[67] M. Dafermos, “Stability and instability of the Cauchy horizon for the spherically symmetric Einstein-Maxwell-scalar field equations”, *Ann. Math. (2)* **158** (2003), no. 3, pp. 875–928.

[68] M. Van de Moortel, “Stability and instability of the sub-extremal Reissner–Nordström black hole interior for the Einstein–Maxwell–Klein–Gordon equations in spherical symmetry”, *Commun. Math. Phys.* **360** (2018), no. 1, pp. 103–168.

[69] Y. Shlapentokh-Rothman, “Exponentially growing finite energy solutions for the Klein–Gordon equation on sub-extremal Kerr spacetimes”, *Commun. Math. Phys.* **329** (2014), no. 3, pp. 859–891.

[70] F. Pasqualotto, Y. Shlapentokh-Rothman and M. Van de Moortel, *The asymptotics of massive fields on stationary spherically symmetric black holes for all angular momenta*, preprint, 2023, 2303.17767.

[71] O. Chodosh and Y. Shlapentokh-Rothman, “Stationary axisymmetric black holes with matter”, *Commun. Anal. Geom.* **29** (2021), no. 1, pp. 19–76.

[72] O. Chodosh and Y. Shlapentokh-Rothman, “Time-periodic Einstein–Klein–Gordon bifurcations of Kerr”, *Commun. Math. Phys.* **356** (2017), no. 3, pp. 1155–1250.

[73] J. Luk and S.-J. Oh, “Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data I. The interior of the black hole region”, *Ann. Math. (2)* **190** (2019), no. 1, pp. 1–111.

[74] M. Dafermos and I. Rodnianski, “A proof of Price’s law for the collapse of a self-gravitating scalar field”, *Invent. Math.* **162** (2005), no. 2, pp. 381–457.

[75] S. Hod and T. Piran, “Late-time evolution of charged gravitational collapse and decay of charged scalar hair. II”, *Phys. Rev. D (3)* **58** (1998), no. 2, article no. 024018.

[76] S. Hod and T. Piran, “Mass inflation in dynamical gravitational collapse of a charged scalar field”, *Phys. Rev. Lett.* **81** (1998), pp. 1554–1557.

[77] S. Klainerman and J. Szeftel, “Kerr stability for small angular momentum”, *Pure Appl. Math. Q.* **19** (2023), no. 3, pp. 791–1678.

[78] E. Giorgi, S. Klainerman and J. Szeftel, *Wave equations estimates and the nonlinear stability of slowly rotating Kerr black holes*, preprint, 2022, 2205.14808. to appear in *Pure Appl. Math. Quarter.*

[79] E. Poisson and W. Israel, “Internal structure of black holes”, *Phys. Rev. D (3)* **41** (1990), no. 6, pp. 1796–1809.

[80] E. Poisson and W. Israel, “Inner-horizon instability and mass inflation in black holes”, *Phys. Rev. Lett.* **63** (1989), no. 16, pp. 1663–1666.

[81] A. Ori, “Inner structure of a charged black hole: an exact mass-inflation solution”, *Phys. Rev. Lett.* **67** (1991), no. 7, pp. 789–792.

[82] W. A. Hiscock, “Evolution of the interior of a charged black hole”, *Phys. Lett. A* **83** (1981), no. 3, pp. 110–112.

[83] M. Van de Moortel, “Mass inflation and the C^2-inextendibility of spherically symmetric charged scalar field dynamical black holes”, *Commun. Math. Phys.* **382** (2021), no. 2, pp. 1263–1341.

[84] J. Luk and S.-J. Oh, “Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data II: the exterior of the black hole region”, *Ann. PDE* **5** (2019), no. 1, article no. 6.

[85] J. Luk, “Weak null singularities in general relativity”, *J. Amer. Math. Soc.* **31** (2018), no. 1, pp. 1–63.

[86] M. Van de Moortel, “Violent nonlinear collapse in the interior of charged hairy black holes”, *Arch. Ration. Mech. Anal.* **248** (2024), article no. 89.

[87] M. Van de Moortel, “The breakdown of weak null singularities inside black holes”, *Duke Math. J.* **172** (2023), no. 15, pp. 2957–3012.

[88] J. Sbierski, "On holonomy singularities in general relativity and the $C^{0,1}_{\mathrm{loc}}$-inextendibility of spacetimes", *Duke Math. J.* **171** (2022), no. 14, pp. 2881–2942.

[89] W. Li and M. Van de Moortel, "Kasner bounces and fluctuating collapse inside hairy black holes with charged matter", *Ann. PDE* **11** (2025), article no. 3.

[90] W. Zheng, *Asymptotically Anti-de-Sitter hairy black holes*, preprint, 2024, 2410.04758.

[91] E. Kasner, "Geometrical theorems on Einstein's cosmological equations", *Amer. J. Math.* **43** (1921), pp. 217–221.

[92] G. Fournodavlos, I. Rodnianski and J. Speck, "Stable big bang formation for Einstein's equations: the complete sub-critical regime", *J. Amer. Math. Soc.* **36** (2023), no. 3, pp. 827–916.

[93] V. A. Belinski, I. M. Khalatnikov and E. M. Lifshitz, "Oscillatory approach to a singular point in relativistic cosmology", *Sov. Phys. Uspekhi* **13** (1971), no. 6, p. 745.

[94] V. A. Belinski, I. M. Khalatnikov and E. M. Lifshitz, "A general solution of the Einstein equations with a time singularity", *Adv. Phys.* **31** (1982), no. 6, pp. 639–667.

[95] M. Van de Moortel, *The coexistence of null and spacelike singularities inside spherically symmetric black holes.* (in preparation).

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

[97] X. An and Z. F. Lim, "Trapped surface formation for spherically symmetric Einstein–Maxwell-charged scalar field system with double null foliation", *Ann. Henri Poincare* **23** (2022), no. 9, pp. 3159–3190.

[98] S. Klainerman, J. Luk and I. Rodnianski, "A fully anisotropic mechanism for formation of trapped surfaces in vacuum", *Invent. Math.* **198** (2014), no. 1, pp. 1–26.

[99] S. Klainerman and I. Rodnianski, "On the formation of trapped surfaces", *Acta Math.* **208** (2012), no. 2, pp. 211–333.

[100] X. An and J. Luk, "Trapped surfaces in vacuum arising dynamically from mild incoming radiation", *Adv. Theor. Math. Phys.* **21** (2017), no. 1, pp. 1–120.

[101] A. Y. Burtscher and P. G. LeFloch, "The formation of trapped surfaces in spherically-symmetric Einstein–Euler spacetimes with bounded variation", *J. Math. Pures Appl. (9)* **102** (2014), no. 6, pp. 1164–1217.

[102] X. An and D. Gajic, "Curvature blow-up and mass inflation in spherically symmetric collapse to a Schwarzschild black hole", *Arch. Ration. Mech. Anal.* **247** (2023), no. 3, article no. 51.

[103] X. An and R. Zhang, "Polynomial blow-up upper bounds for the Einstein-scalar field system under spherical symmetry", *Commun. Math. Phys.* **376** (2020), no. 2, pp. 1671–1704.

[104] W. Li, "Kasner-like description of spacelike singularities in spherically symmetric spacetimes with scalar matter", *Ann. PDE* **11** (2025), article no. 5.

[105] X. An and H. K. Tan, *A proof of weak cosmic censorship conjecture for the spherically symmetric Einstein–Maxwell–Charged scalar field system*, preprint, 2024, 2402.16250.

[106] C. Kehle and R. Unger, "Gravitational collapse to extremal black holes and the third law of black hole thermodynamics", *J. Eur. Math. Soc.* (2025). (Online first).

[107] G. Fournodavlos and J. Luk, "Asymptotically Kasner-like singularities", *Amer. J. Math.* **145** (2023), no. 4, pp. 1183–1272.

[108] H. Ringström, "The Bianchi IX attractor", *Ann. Henri Poincaré* **2** (2001), no. 3, pp. 405–500.

[109] R. H. Price, "Nonspherical perturbations of relativistic gravitational collapse. I. Scalar and gravitational perturbations", *Phys. Rev. D (3)* **5** (1972), pp. 2419–2438.

[110] P. Blue and A. Soffer, "Phase space analysis on some black hole manifolds", *J. Funct. Anal.* **256** (2009), no. 1, pp. 1–90.

[111] M. Dafermos and I. Rodnianski, "The red-shift effect and radiation decay on black hole spacetimes", *Commun. Pure Appl. Math.* **62** (2009), no. 7, pp. 859–919.

[112] L. Andersson, T. Bäckdahl, P. Blue and S. Ma, *Stability for linearized gravity on the Kerr spacetime*, preprint, 2019, 1903.03859.

[113] D. Tataru, "Local decay of waves on asymptotically flat stationary space-times", *Amer. J. Math.* **130** (2008), no. 3, pp. 571–634.

[114] L. Andersson and P. Blue, "Hidden symmetries and decay for the wave equation on the Kerr spacetime", *Ann. Math. (2)* **182** (2015), no. 3, pp. 787–853.

[115] R. Donninger, W. Schlag and A. Soffer, "On pointwise decay of linear waves on a schwarzschild black hole background", *Commun. Math. Phys.* **309** (2012), pp. 51–86.

[116] R. Donninger, W. Schlag and A. Soffer, "A proof of Price's Law on Schwarzschild black hole manifolds for all angular momenta", *Adv. Math.* **226** (2011), no. 1, pp. 484–540.

[117] J. Metcalfe, D. Tataru and M. Tohaneanu, "Price's law on nonstationary space-times", *Adv. Math.* **230** (2012), no. 3, pp. 995–1028.

[118] P. Hintz, "A sharp version of Price's law for wave decay on asymptotically flat spacetimes", *Commun. Math. Phys.* **389** (2022), pp. 491–542.

[119] Y. Angelopoulos, S. Aretakis and D. Gajic, "Late-time asymptotics for the wave equation on spherically symmetric, stationary spacetimes", *Adv. Math.* **323** (2018), pp. 529–621.

[120] Y. Angelopoulos, S. Aretakis and D. Gajic, "Price's Law and precise late-time asymptotics for subextremal reissner–nordström black holes", *Ann. Henri Poincare* **24** (2023), no. 9, pp. 3215–3287.

[121] S. A. Teukolsky, "Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations", *Astrophys. J.* **185** (1973), pp. 635–647.

[122] Y. Shlapentokh-Rothman and R. Teixeira da Costa, *Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: frequency space analysis*, preprint, 2020, 2007.07211.

[123] Y. Shlapentokh-Rothman and R. Teixeira da Costa, *Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: physical space analysis*, preprint, 2023, 2302.08916.

[124] M. Dafermos, G. Holzegel and I. Rodnianski, "Boundedness and decay for the Teukolsky equation on Kerr spacetimes I: The case $|a| \ll M$", *Ann. PDE* **5** (2019), no. 1, article no. 2.

[125] S. Ma, "Uniform energy bound and Morawetz estimate for extreme components of spin fields in the exterior of a slowly rotating Kerr black hole II: Linearized gravity", *Commun. Math. Phys.* **377** (2020), no. 3, pp. 2489–2551.

[126] L. Andersson and P. Blue, "Uniform energy bound and asymptotics for the Maxwell field on a slowly rotating Kerr black hole exterior", *J. Hyperbolic Differ. Equ.* **12** (2015), no. 4, pp. 689–743.

[127] P. Millet, *Optimal decay for solutions of the Teukolsky equation on the Kerr metric for the full subextremal range $|a| < M$*, preprint, 2023, 2302.06946.

[128] S. Ma, "Almost Price's law in Schwarzschild and decay estimates in Kerr for Maxwell field", *J. Differ. Equ.* **339** (2022), pp. 1–89.

[129] S. Ma and L. Zhang, "Sharp decay for Teukolsky equation in Kerr spacetimes", *Commun. Math. Phys.* **401** (2023), pp. 333–434.

[130] S. Ma and L. Zhang, "Sharp decay estimates for massless Dirac fields on a Schwarzschild background", *J. Funct. Anal.* **282** (2022), no. 6, article no. 109375.

[131] F. Pasqualotto, "The spin ± 1 Teukolsky equations and the Maxwell system on Schwarzschild", *Ann. Henri Poincaré* **20** (2019), no. 4, pp. 1263–1323.

[132] J. Sterbenz and D. Tataru, "Local energy decay for Maxwell fields part I: Spherically symmetric black-hole backgrounds", *Int. Math. Res. Not.* **2015** (2015), no. 11, pp. 3298–3342.

[133] J. Metcalfe, D. Tataru and M. Tohaneanu, "Pointwise decay for the Maxwell field on black hole space-times", *Adv. Math.* **316** (2017), pp. 53–93.

[134] P. Blue, "Decay of the Maxwell field on the Schwarzschild manifold", *J. Hyperbolic Differ. Equ.* **5** (2008), no. 4, pp. 807–856.

[135] E. Giorgi, "The linear stability of Reissner–Nordström spacetime: the full subextremal range $|Q| < M$", *Commun. Math. Phys.* **380** (2020), no. 3, pp. 1313–1360.

[136] M. Van de Moortel, "Decay of weakly charged solutions for the spherically symmetric Maxwell-charged-scalar-field equations on a Reissner–Nordström exterior space-time", *Ann. Sci. Éc. Norm. Supér. (4)* **55** (2022), no. 2, pp. 283–404.

[137] D. Gajic and M. Van de Moortel, *Late-time tails for scale-invariant wave equations with a potential and the near-horizon geometry of null infinity*, preprint, 2024, 2401.13047.

[138] S. Detweiler, "Klein–Gordon equation and rotating black holes", *Phys. Rev. D* **22** (1980), pp. 2323–2326.

[139] T. J. Zouros and D. M. Eardley, "Instabilities of massive scalar perturbations of a rotating black hole", *Ann. Phys.* **118** (1979), no. 1, pp. 139–155.

[140] R. Brito, V. Cardoso and P. Pani, *Superradiance—New Frontiers in Black Hole Physics*, Second edition, Lecture Notes in Physics, Springer, 2020.

[141] F. C. Eperon, B. Ganchev and J. E. Santos, "Plausible scenario for a generic violation of the weak cosmic censorship conjecture in asymptotically flat four dimensions", *Phys. Rev. D* **101** (2020), no. 4, article no. 041502(R).

[142] B. Ganchev and J. E. Santos, "Scalar hairy black holes in four dimensions are unstable", *Phys. Rev. Lett.* **120** (2018), article no. 171101.

[143] Y. Shlapentokh-Rothman and M. Van de Moortel, *Polynomial time decay for solutions of the Klein–Gordon equation on a subextremal Reissner–Nordström black hole*, preprint, 2024, 2401.00048.

[144] J. Luk and S.-J. Oh, *Late time tail of waves on dynamic asymptotically flat spacetimes of odd space dimensions*, preprint, 2024, 2404.02220.

[145] E. Giorgi, S. Klainerman and J. Szeftel, *A general formalism for the stability of Kerr*, preprint, 2020, 2002.02740.

[146] S. Klainerman and J. Szeftel, "Construction of GCM spheres in perturbations of Kerr", *Ann. PDE* **8** (2022), no. 2, article no. 17.

[147] S. Klainerman and J. Szeftel, "Effective results on uniformization and intrinsic GCM spheres in perturbations of Kerr", *Ann. PDE* **8** (2022), no. 2, article no. 18.

[148] D. Shen, "Construction of GCM hypersurfaces in perturbations of Kerr", *Ann. PDE* **9** (2023), no. 1, article no. 11.

[149] M. Dafermos, G. Holzegel, I. Rodnianski and M. Taylor, *The non-linear stability of the Schwarzschild family of black holes*, preprint, 2021, 2104.08222.

[150] M. Dafermos, G. Holzegel and I. Rodnianski, *Quasilinear wave equations on asymptotically flat spacetimes with applications to Kerr black holes*, preprint, 2022, 2212.14093.

[151] S. Klainerman and J. Szeftel, *Global Nonlinear Stability of Schwarzschild Spacetime Under Polarized Perturbations*, Annals of Mathematics Studies, Princeton University Press: Princeton, NJ, 2020.

[152] D. Häfner, P. Hintz and A. Vasy, "Linear stability of slowly rotating Kerr black holes", *Invent. Math.* **223** (2021), no. 3, pp. 1227–1406.

[153] C. Kehle and R. Unger, *Extremal black hole formation as a critical phenomenon*, preprint, 2024, 2402.10190.

[154] D. Gajic and J. Luk, "The interior of dynamical extremal black holes in spherical symmetry", *Pure Appl. Anal.* **1** (2019), no. 2, pp. 263–326.

[155] D. Gajic, "Linear waves in the interior of extremal black holes I", *Commun. Math. Phys.* **353** (2017), no. 2, pp. 717–770.

[156] D. Gajic, "Linear waves in the interior of extremal black holes II", *Ann. Henri Poincaré* **18** (2017), no. 12, pp. 4005–4081.

[157] M. Van de Moortel, *Charged scalar fields on black hole space-times*, PhD thesis, University of Cambridge, 2019.

[158] S. Hawking, "The occurrence of singularities in cosmology. III. Causality and singularities", *Proc. R. Soc. Lond. A* **300** (1967), pp. 187–201.

[159] V. Moncrief and J. Isenberg, "Symmetries of cosmological Cauchy horizons with non-closed orbits", *Commun. Math. Phys.* **374** (2020), no. 1, pp. 145–186.

[160] O. Lindblad Petersen, "Extension of Killing vector fields beyond compact Cauchy horizons", *Adv. Math.* **391** (2021), article no. 107953.

[161] H. Friedrich, I. Rácz and R. M. Wald, "On the rigidity theorem for spacetimes with a stationary event horizon or a compact Cauchy horizon", *Commun. Math. Phys.* **204** (1999), no. 3, pp. 691–707.

[162] R. H. Gowdy, "Vacuum spacetimes with two-parameter spacelike isometry groups and compact invariant hypersurfaces: topologies and boundary conditions", *Ann. Phys.* **83** (1974), pp. 203–241.

[163] P. T. Chruściel, "On space-times with U(1) × U(1) symmetric compact Cauchy surfaces", *Ann. Phys.* **202** (1990), no. 1, pp. 100–150.

[164] H. Ringstrom, "Cosmic censorship for gowdy spacetimes", *Living Rev. Rel.* **13** (2010), article no. 2.

[165] H. Ringström, "Strong cosmic censorship in T^3-Gowdy spacetimes", *Ann. Math. (2)* **170** (2009), no. 3, pp. 1181–1240.

[166] H. Ringström, "Existence of an asymptotic velocity and implications for the asymptotic behavior in the direction of the singularity in T^3-Gowdy", *Commun. Pure Appl. Math.* **59** (2006), no. 7, pp. 977–1041.

[167] M. Dafermos and A. D. Rendall, "Inextendibility of expanding cosmological models with symmetry", *Class. Quantum Gravity* **22** (2005), no. 23, pp. L143–L147.

[168] M. Dafermos and A. D. Rendall, "An extension principle for the Einstein–Vlasov system in spherical symmetry", *Ann. Henri Poincaré* **6** (2005), no. 6, pp. 1137–1155.

[169] M. Dafermos and A. D. Rendall, "Strong cosmic censorship for surface-symmetric cosmological spacetimes with collisionless matter", *Commun. Pure Appl. Math.* **69** (2016), no. 5, pp. 815–908.

[170] J. Smulevici, "Strong cosmic censorship for T^2-symmetric spacetimes with cosmological constant and matter", *Ann. Henri Poincaré* **9** (2008), no. 8, pp. 1425–1453.

[171] C. W. Misner, "Mixmaster universe", *Phys. Rev. Lett.* **22** (1969), pp. 1071–1074.

[172] H. Oude Groeniger, O. Petersen and H. Ringström, *Formation of quiescent big bang singularities*, preprint, 2023, 2309.11370.

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

[174] S. C. Collingbourne and G. Holzegel, "Uniform boundedness for solutions to the Teukolsky equation on Schwarzschild from conservation laws of linearised gravity", *Commun. Math. Phys.* **405** (2024), no. 6, article no. 138.

[175] S. C. Collingbourne, "Coercivity properties of the canonical energy in double null gauge on the 4-dimensional Schwarzschild exterior", *Class. Quant. Grav.* **40** (2023), no. 22, article no. 225013.

[176] P. Laul, J. Metcalfe, S. Tikare and M. Tohaneanu, "Localized energy estimates for wave equations on (1 + 4)-dimensional Myers-Perry space-times", *SIAM J. Math. Anal.* **47** (2015), no. 3, pp. 1933–1957.

[177] P.-K. Hung, J. Keller and M.-T. Wang, "Linear stability of higher dimensional Schwarzschild spacetimes: decay of master quantities", *Ann. PDE* **6** (2020), no. 2, article no. 7.

[178] V. Schlue, "Decay of linear waves on higher-dimensional Schwarzschild black holes", *Anal. PDE* **6** (2013), no. 3, pp. 515–600.

[179] R. Emparan and H. S. Reall, "Black rings", *Class. Quantum Gravity* **23** (2006), no. 20, R169–R197.

[180] R. Gregory and R. Laflamme, "Black strings and p-branes are unstable", *Phys. Rev. Lett.* **70** (1993), pp. 2837–2840.

[181] S. C. Collingbourne, "The Gregory–Laflamme instability of the Schwarzschild black string exterior", *J. Math. Phys.* **62** (2021), no. 3, article no. 032502.

[182] G. Benomio, "The stable trapping phenomenon for black strings and black rings and its obstructions on the decay of linear waves", *Anal. PDE* **14** (2021), no. 8, pp. 2427–2496.

[183] A. J. Fang, *Linear stability of the slowly-rotating Kerr-de Sitter family*, preprint, 2022, 2207.07902.

[184] G. Mavrogiannis, "Morawetz estimates without relative degeneration and exponential decay on Schwarzschild–de Sitter spacetimes", *Ann. Henri Poincaré* **24** (2023), no. 9, pp. 3113–3152.

[185] S. Dyatlov, "Quasi-normal modes and exponential energy decay for the Kerr–de Sitter black hole", *Commun. Math. Phys.* **306** (2011), no. 1, pp. 119–163.

[186] A. Vasy, "Microlocal analysis of asymptotically hyperbolic and Kerr-de Sitter spaces (with an appendix by Semyon Dyatlov)", *Invent. Math.* **194** (2013), no. 2, pp. 381–513.

[187] P. Hintz and A. Vasy, "Global analysis of quasilinear wave equations on asymptotically Kerr–de Sitter spaces", *Int. Math. Res. Not.* **2016** (2016), no. 17, pp. 5355–5426.

[188] G. Mavrogiannis, "Quasilinear wave equations on Schwarzschild–de Sitter", *Commun. Partial Differ. Equ.* **49** (2024), no. 1-2, pp. 38–87.

[189] P. Hintz and A. Vasy, "Semilinear wave equations on asymptotically de Sitter, Kerr–de Sitter and Minkowski spacetimes", *Anal. PDE* **8** (2015), no. 8, pp. 1807–1890.

[190] P. Hintz and A. Vasy, "The global non-linear stability of the Kerr–de Sitter family of black holes", *Acta Math.* **220** (2018), no. 1, pp. 1–206.

[191] A. J. Fang, *Nonlinear stability of the slowly-rotating Kerr-de Sitter family*, preprint, 2021, 2112.07183.

[192] P. Hintz and A. Vasy, "Analysis of linear waves near the Cauchy horizon of cosmological black holes", *J. Math. Phys.* **58** (2017), no. 8, article no. 081509.

[193] V. Cardoso, J. a. L. Costa, K. Destounis, P. Hintz and A. Jansen, "Quasinormal modes and strong cosmic censorship", *Phys. Rev. Lett.* **120** (2018), article no. 031103.

[194] M. Casals and C. I. S. Marinho, "Glimpses of violation of strong cosmic censorship in rotating black holes", *Phys. Rev. D* **106** (2022), article no. 044060.

[195] O. J. C. Dias, F. C. Eperon, H. S. Reall and J. E. Santos, "Strong cosmic censorship in de Sitter space", *Phys. Rev. D* **97** (2018), article no. 104060.

[196] F. Rossetti, "Strong cosmic censorship for the spherically symmetric Einstein–Maxwell–Charged–Klein–Gordon system with positive Λ: stability of the Cauchy horizon and H^1 Extensions", *Ann. Henri Poincaré* (2024). to appear.

[197] J. L. Costa, P. M. Girão, J. Natário and J. D. Silva, "On the global uniqueness for the Einstein–Maxwell–Scalar field system with a cosmological constant: Part 3. Mass inflation and extendibility of the solutions", *Ann. PDE* **3** (2017), article no. 8.

[198] J. L. Costa, P. M. Girão, J. Natário and J. D. Silva, "On the occurrence of mass inflation for the Einstein–Maxwell-Scalar field system with a cosmological constant and an exponential price law", *Commun. Math. Phys.* **361** (2018), no. 1, pp. 289–341.

[199] G. Holzegel and J. Smulevici, "Quasimodes and a lower bound on the uniform energy decay rate for Kerr-AdS spacetimes", *Anal. PDE* **7** (2014), no. 5, pp. 1057–1090.

[200] O. Gannot, "Quasinormal modes for Schwarzschild-AdS black holes: exponential convergence to the real axis", *Commun. Math. Phys.* **330** (2014), no. 2, pp. 771–799.

[201] W. Zheng, *Exponentially-growing Mode Instability on Reissner–Nordström-Anti-de-Sitter black holes*, preprint, 2024, 2410.04750.

[202] S. A. Hartnoll, C. P. Herzog and G. T. Horowitz, "Holographic superconductors", *J. High Energy Phys.* **2008** (2008), no. 12, article no. 015.

[203] O. J. C. Dias and R. Masachs, "Hairy black holes and the endpoint of AdS_4 charged superradiance", *J. High Energy Phys.* (2017), no. 2, article no. 128.

[204] C. Kehle, "Uniform boundedness and continuity at the Cauchy horizon for linear waves on Reissner–Nordström-AdS black holes", *Commun. Math. Phys.* **376** (2020), no. 1, pp. 145–200.

[205] C. Kehle, "Blowup of the local energy of linear waves at the Reissner–Nordström-AdS Cauchy horizon", *Class. Quantum Gravity* **38** (2021), no. 21, article no. 214001.

[206] C. Kehle, "Diophantine approximation as cosmic censor for Kerr-AdS black holes", *Invent. Math.* **227** (2022), no. 3, pp. 1169–1321.

[207] P. Figueras and L. Rossi, *Non-linear instability of slowly rotating Kerr-AdS black holes*, preprint, 2023, 2311.14167.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 53-78
https://doi.org/10.5802/crmeca.277

Review article / *Article de synthèse*

Cosmology, the big bang and the BKL conjecture

Cosmologie, big bang et conjecture BKL

Hans Ringström [a]

[a] Department of Mathematics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
E-mail: hansr@kth.se

Abstract. This is a review article of mathematical results in cosmology, written in honor of Yvonne Choquet-Bruhat's 100th birthday. It starts with a brief description of some of the essential questions: strong cosmic censorship; the relation between the future asymptotics and geometrization in the vacuum setting; the cosmic no-hair conjecture; and the BKL-proposal. It then turns to results, starting with ones obtained in situations with symmetry. It continues with a review of future global non-linear stability results, stable big bang formation results, results concerning solutions to linear equations on cosmological backgrounds and numerical results. The article contains a substantial, but very incomplete, list of references to the literature.

Résumé. Il s'agit d'un article de revue sur les résultats mathématiques en cosmologie, écrit en l'honneur du 100ème anniversaire d'Yvonne Choquet-Bruhat. Il commence par une brève description de certaines questions essentielles : la censure cosmique forte, la relation entre l'asymptotique future et la géométrisation dans le cas du vide, la conjecture de l'absence de cheveux dans le cadre cosmique et la proposition BKL. Il aborde ensuite les résultats, en commençant par ceux obtenus dans des situations avec symétrie. Il se poursuit par un examen des résultats de stabilité future non linéaire globale, des résultats de formation stable du big bang, des résultats concernant les solutions d'équations linéaires sur les espaces-temps cosmologiques et des résultats numériques. L'article contient une liste substantielle, mais très incomplète, de références à la littérature.

Keywords. Global analysis, General relativity, Lorentzian geometry, Nonlinear evolution equations, Cosmology, The big bang and the BKL proposal.

Mots-clés. Analyse globale, Relativité générale, Géométrie lorentzienne, Équations d'évolution non linéaires, Cosmologie, Big bang et proposition BKL.

Funding. Vetenskapsrådet, the Swedish Research Council, dnr. 2022-03053.

Manuscript received 11 October 2024, revised 2 December 2024, accepted 3 December 2024.

1. Introduction

From Einstein's attempt to find a static model of the universe; to Friedmann's discovery of expanding solutions; Lemaître's conclusion that the universe is expanding, based on a combination of observations and expanding solutions; and Hubble's observations, the initial history of general relativistic cosmology was long and complicated, and we refer the interested reader to [1] for a detailed discussion. Once it becomes clear that the universe is expanding, it is natural to trace that expansion backward in order to find an origin. However, even if one finds models with a corresponding cosmological, or "big bang", singularity, it is not clear if that is the only

explanation, nor is it clear that such a singularity would persist when relaxing the symmetry assumptions associated with the model. Hawking's theorem gives one indication that cosmological singularities are a general feature of solutions to Einstein's equations; see [2, 3] for original references and [4, pp. 431–434] for a textbook presentation. Moreover, the discovery of the cosmic microwave background (CMB) radiation by Penzias and Wilson [5] gives clear observational corroboration to the idea of a big bang. However, in spite of the importance of Hawking's result, it does not say much about the nature of singularities. Does the gravitational field generically blow up in the direction of the singularity? Is the spacetime inextendible? In order to answer these questions, it would seem to be necessary to carry out a more detailed analysis of solutions. Moreover, the high degree of isotropy of the CMB cannot be explained by a thermal equilibrium having been established prior to the surface of last scattering due to the causal structure of the standard models of the universe. For this reason, either more sophisticated models of the universe or more exotic models of matter (inducing, e.g., inflation) would seem to be necessary.

Already at this point, the necessity for a deeper understanding of the space of cosmological solutions to Einstein's equations is apparent. This can be viewed as a mathematical problem, and the purpose of this review article is to give a rough idea of the developments in this subject since the appearance of Yvonne Choquet-Bruhat's seminal paper [6]. Due to the number of results that have been obtained, giving a complete description is not possible. In fact, this review article describes a small subset of the results, and the limited selection is due to the knowledge (or absence thereof) and interests of the author. We begin the discussion in the next section by highlighting some of the questions. In the following sections, we describe results.

2. Cosmology, conjectures

Before describing the objects of study, it is natural to note that there are at least two fundamental problems in the subject of cosmology: we cannot carry out experiments; and we only observe a small part of the universe. Deducing the global properties of the universe on the basis of observations might therefore even in principle be impossible. In cosmology, it is common to compensate for the lack of knowledge by imposing philosophical principles. One example is the *Copernican principle*, a strong version of which states that there are no privileged observers in the universe. This principle does not require the universe to appear the same to all observers. Nevertheless, the currently preferred so called *cosmological principle*, states that the universe can be modeled by a spatially homogeneous and isotropic solution. In such solutions, one cannot distinguish between two points in space, nor between two directions. This quite extreme assumption reduces the possibilities for the geometry to a choice of a 3-dimensional Riemannian manifold of constant curvature and a scalar function of time, describing the expansion and contraction of space. Physicists largely restrict themselves to this situation and are more likely, when problems arise, to introduce exotic matter models than to consider more general geometries. In the mathematical study of cosmology, we do want to consider more general geometries. On the other hand, we are not interested in, e.g., asymptotically flat solutions, in which there are clearly preferred asymptotic regimes. Using the fact that there is a well posed Cauchy problem for Einstein's equations, see e.g. [6–12], it is natural to describe the conditions in terms of initial data. We do not attempt to give a formal definition of the meaning of a cosmological spacetime here (where by a *spacetime*, we mean a time oriented Lorentz manifold). However, we have a spacetime in mind in which the geometry and matter fields in different parts of the initial hypersurface are qualitatively similar. In particular, initial data on a closed manifold such that the mean curvature is negative on some parts and positive on others is not of the type we wish to consider here; this would mean that some observers experience expansion and some contraction. In particular, that the initial manifold is closed is neither a necessary nor a sufficient condition for the corresponding spacetime to

be cosmological. In fact, we are mainly interested in initial data sets for which the mean curvature is bounded away from zero by a non-zero constant, in analogy with the conditions of Hawking's theorem.

Next, we describe some of the questions of interest.

2.1. *Strong cosmic censorship, curvature blow up*

As mentioned in the introduction, Hawking's singularity theorem, see e.g. [4, Theorem 55A, p. 431], ensures that singularities appear naturally in general relativity. On the other hand, the only conclusion provided is that there are incomplete timelike geodesics. Moreover, letting N be the timelike future of a point in Minkowski space, letting g denote the Minkowski metric restricted to N, and foliating (N, g) with hyperboloids, the conditions of Hawking's theorem are satisfied (and indicate the occurrence of a singularity), even though (N, g) is simply a subset of Minkowski space. Clearly, it is desirable to obtain more information. At the very least, it would be desirable to prove that if one traces a causal geodesic in an incomplete direction, then a curvature invariant, such as the Kretschmann scalar (i.e., the Riemann curvature tensor contracted with itself), becomes unbounded. Unfortunately, such a statement cannot hold for arbitrary initial data. However, one can conjecture that for generic initial data in an appropriately chosen class, the Kretschmann scalar is unbounded in the incomplete directions of causal geodesics in the corresponding maximal Cauchy development; this is the *curvature blow up conjecture*. It is natural to try to prove the corresponding statement in symmetry classes of solutions. This conjecture is related to the *strong cosmic censorship conjecture*, stating that for generic initial data (in an appropriate class), the corresponding maximal Cauchy development is inextendible. This would mean that, even though Einstein's equations are not deterministic (in the sense that one cannot determine what spacetime one is in on the basis of initial data) they are generically deterministic. In order to obtain mathematical results, one of course has to give a precise definition of the class of data considered and the meaning of the word inextentible.

2.2. *Vacuum evolution and geometrization*

In [13], Eardley and Moncrief suggested one way to address the strong cosmic censorship conjecture: by proving global existence results for constant mean curvature (CMC) foliations. Their hope was that the CMC foliations should avoid the singularities and cover the spacetimes outside the singularities. More specifically, and focusing on the expanding direction, there are, due to the work of Fischer and Moncrief on the one hand, see [14–22], and Anderson on the other, see [23, 24], quite detailed conjectures concerning the future asymptotics of solutions to Einstein's vacuum equations (with a vanishing cosmological constant) in the cosmological setting. In these conjectures, the initial manifold is assumed to be closed and to not admit a Riemannian metric of positive scalar curvature. Moreover, the authors focus on CMC foliations. Since the volume of the leaves of the CMC foliation is expected to tend to infinity in the expanding direction, it is natural to carry out an expansion normalization of the metric. In the work of Fischer and Moncrief, the authors rescale the Riemannian metrics on the leaves by multiplying them by the mean curvature squared. Anderson rescales by dividing by the proper time squared (where the proper time is measured to a fixed Cauchy hypersurface). Considering the rescaled metrics as a family of Riemannian metrics on a fixed closed manifold, the idea is that in the limit, this family implements a geometrization of the manifold. In particular, it converges to complete hyperbolic metrics of finite volume on the hyperbolic pieces and collapses on the Seifert fibred pieces. In this sense the evolution leads to an isotropization, on average, in the expanding direction. As will become clear below, there are many results confirming this picture. However, the results are

either in symmetric settings or concern perturbations of symmetric settings. This means that the initial manifold must admit a symmetric metric, so that the geometrization must be trivial. In other words, confirming the conjectured picture in a non-trivial setting can be expected to be quite difficult.

2.3. *Cosmic no-hair conjecture*

Turning to cosmological solutions with a positive cosmological constant, the *cosmic no-hair conjecture* states that, in the expanding direction, spacetimes should asymptote to de Sitter space. More specifically, fixing, e.g., a future complete timelike geodesic, say γ, then the behavior in $J^-(\gamma)$ should asymptotically be indistinguishable from de Sitter space. For a more formal definition, see [25, Definition 1.8, pp. 1573–1574] and [25, Conjecture 1.11, p. 1574]. Again, there are many results confirming this proposal in specific situations.

2.4. *The BKL proposal*

This review article is focused on mathematical results. However, the questions that are addressed in the mathematical literature are at least partially inspired by heuristic arguments and proposals arising in the physics literature. One such proposal is due to Belinskiĭ, Khalatnikov and Lifschitz, BKL for short, and dates back to the 60's, 70's and 80's; see, e.g., [26–31]. The rough statement of the BKL proposal is that generic cosmological singularities are spacelike, local and oscillatory (with the exception of matter models such as scalar fields and stiff fluids). More specifically, distinct causal curves typically lose the ability to communicate in the direction of the singularity; each causal curve evolves as its own universe in the direction of the singularity; and the local behavior is modeled by the dynamics of Bianchi type VIII or IX (mixmaster) spactimes, which are expected to exhibit (chaotic) oscillatory behavior in the direction of the singularity. A related idea arising from the BKL proposal is that spatial derivatives should be dominated by time derivatives in the approach to the singularity. From this perspective, it is natural to drop the spatial derivatives in the equations, which leads to an interpretation of Einstein's equations as a family of ODE's (one ODE for each spatial point). Two examples of attempts to make this perspective more precise are given by [32, 33]. See also Section 2.5 below for a more detailed discussion. The ideas of BKL have later been developed further in the physics literature, see, e.g., [34–37]. More mathematical results, deriving conclusions from a priori assumptions, can be found in, e.g., [38–40]. In the next subsection, we try to make the ideas of BKL more precise in the non-oscillatory setting.

2.5. *Initial data on the singularity*

One of the ideas of BKL is that the dynamics of nearby observers decouple and that time derivative terms (or "kinetic terms") asymptotically dominate over spatial derivative terms (or spatial curvature terms) in the direction of the singularity. In [32], the authors discuss ideas of this type in terms of asymptotics of the geometry, and they speak of *velocity-dominated singularities*. A related perspective is developed in [33], in which the authors introduce the terminology that a spacetime can be *asymptotically velocity term dominated (AVTD) near the singularity*. The idea in [33] is to introduce a truncated set of equations, the *velocity term dominated (VTD) system*, and to say that a solution is AVTD if it asymptotes to a solution to the VTD system; see [33] for details. Unfortunately, the VTD system has no covariant formulation. On the other hand, the evolutionary part of the VTD system consists of ODE's for each spatial point. This means that the VTD system is significantly easier to solve than the original equations. Moreover, in [41] the

authors prove that, given a real analytic solution to the VTD system in the $3+1$-dimensional Einstein-scalar field or Einstein-stiff fluid setting (satisfying certain conditions), there is a unique corresponding solution to the actual equations. The authors of [41] call the resulting singularities *quiescent*, a terminology that goes back to [42]. In the $3+1$-dimensional vacuum setting, the behavior in the direction of the singularity is expected to be oscillatory (in accordance with the BKL proposal), in which case it is, at this time, not clear how to specify initial data on the singularity. However, in higher dimensions (more specifically, for $n+1$-dimensions with $n \geq 10$), quiescence is expected even in the vacuum setting; see [43]. Accordingly, [41] is generalized in [44] to higher dimensions and more general matter fields.

The solutions constructed in [41] and [44] are expected to be robust. In other words, perturbations of the corresponding initial data are expected to lead to maximal Cauchy developments with similar singularities; see Section 5 below for a partial justification of this expectation. However, in situations with symmetry, e.g., even singularities in $3+1$-dimensional vacuum spacetimes can be quiescent; see, e.g, [45–53]. The reason for this is that the symmetry type can eliminate the mechanism causing the oscillations. One can also impose conditions on data on the singularity that eliminate this mechanism. One corresponding result in the real analytic setting, but in the absence of symmetries, is obtained in [54]. A more general result in the smooth setting is contained in [55].

The VTD solutions in [41, 44] can be interpreted as initial data on the singularity. However, there are many other notions of initial data on the singularity arising when studying symmetry classes of solutions as well as in results such as [54, 55]. For this reason, it is of interest to try to find a unifying perspective. This is the goal of [56]. In particular, [56, Definition 10, p. 5] yields the following definition of initial data on the singularity in the Einstein-scalar field setting with a cosmological constant:

Definition 1. *Let $3 \leq n \in \mathbb{N}$, $\Lambda \in \mathbb{R}$, $(\Sigma, \mathscr{H})$ be a smooth n-dimensional Riemannian manifold, $\mathscr{K}$ be a smooth $(1,1)$-tensor field on Σ and Φ_0 and Φ_1 be smooth functions on Σ. Then $(\Sigma, \mathscr{H}, \mathscr{K}, \Phi_0, \Phi_1)$ are non-degenerate quiescent initial data on the singularity for the Einstein-scalar field equations with a cosmological constant Λ if*

(1) $\mathrm{tr}\mathscr{K} = 1$ *and $\mathscr{K}$ is symmetric with respect to $\mathscr{H}$.*
(2) $\mathrm{tr}\mathscr{K}^2 + \Phi_1^2 = 1$ *and* $\mathrm{div}_{\mathscr{H}}\mathscr{K} = \Phi_1 d\Phi_0$.
(3) *The eigenvalues of $\mathscr{K}$, say $\{p_A\}$, are distinct.*
(4) *$\mathscr{O}^A_{BC}$ vanishes in a neighborhood of $\bar{x} \in \Sigma$ if $1 + p_A(\bar{x}) - p_B(\bar{x}) - p_C(\bar{x}) \leq 0$, where $\mathscr{O}^A_{BC} := (\gamma^A_{BC})^2$; γ^A_{BC} is given by $[\mathscr{X}_B, \mathscr{X}_C] = \gamma^A_{BC}\mathscr{X}_A$; and $\mathscr{X}_A$ is an eigenvector field corresponding to p_A such that $|\mathscr{X}_A|_{\mathscr{H}} = 1$.*

The reader interested in a justification of why this is a natural notion of initial data on the singularity is referred to [40, 56]. That the solutions to the VTD equations that are used in [41] as a substitute for data on the singularity are equivalent to data on the singularity in the sense of Definition 1 is demonstrated in [56, Section 1.5, pp. 7–9]. That the notion of initial data used in [55] is a special case of Definition 1 is demonstrated in [56, Section 1.5, pp. 7–9]. That the notion of initial data on the singularity used in [45, 47] in the case of $\mathbb{T}^3$-Gowdy is a special case of Definition 1 is demonstrated in [57]. It would be of interest to show that all the notions of initial data on the singularity used in the literature (in the Einstein-scalar field setting) are special cases of Definition 1. In [58], Andrés Franco Grisales demonstrates that data as in Definition 1 give rise to unique developments in the $3+1$-dimensional Einstein-non-linear scalar field setting. In particular, ignoring the condition of non-degeneracy (i.e., the requirement that the eigenvalues be distinct), [58] simultaneously generalizes [41] and [55].

In $3+1$-dimensions, one way to avoid the condition of non-degeneracy in Definition 1 is given in [56, Remark 26, p. 9]. It is also of great interest to generalize Definition 1 to other matter fields.

Partly, this has already been done in [44]; the VTD solutions in [44] can presumably be used to this end. However, there are many matter models of interest that are not covered in [44].

The relation between solutions and data on the singularity are the following. Assume (M, g, ϕ) to be a solution to the Einstein-scalar field equations with a cosmological constant Λ. Assume, moreover, that there is a partial foliation of (M, g, ϕ) by spacelike Cauchy hypersurfaces, say $\{\Sigma_t\}$ for $t \in (0, t_+)$ and assume that the mean curvature of Σ_t, say θ_t is such that θ_t converges uniformly to ∞ as $t \downarrow 0$. Define $\mathscr{K}_t$ to be the *expansion normalized Weingarten map* of Σ_t, i.e., the Weingarten map of Σ_t divided by θ_t. Define $\mathscr{H}_t := \bar{g}_t(\theta_t^{\mathscr{K}_t}\cdot, \theta_t^{\mathscr{K}_t}\cdot)$, where $\bar{g}_t$ is the Riemannian metric induced on Σ_t. Note that $\mathscr{K}_t$ is an operator on the tangent space of Σ_t and that θ_t can be assumed to be strictly positive, so that

$$\theta_t^{\mathscr{K}_t} := \sum_{l=0}^{\infty} \frac{(\ln\theta_t)^l \mathscr{K}_t^l}{l!}.$$

Finally, let ϕ_t be the scalar field induced on Σ_t, $\dot{\phi}_t$ be the future unit normal derivative of ϕ at Σ_t, $\Phi_{1,t} := \dot{\phi}_t/\theta_t$ and $\Phi_{0,t} := \phi_t + \Phi_{1,t}\ln\theta_t$. Then (M, g, ϕ) is said to be a *development of initial data* $(\Sigma, \mathscr{H}, \mathscr{K}, \Phi_0, \Phi_1)$ *on the singularity* if

$$\mathscr{H}_t \to \mathscr{H}, \quad \mathscr{K}_t \to \mathscr{K}, \quad \Phi_{0,t} \to \Phi_0, \quad \Phi_{1,t} \to \Phi_1 \tag{1}$$

as $t \downarrow 0$, where the topology defining the convergence should be adapted to the situation.

The notion of initial data on the singularity serves many purposes. First, the definition is covariant. Second, that a solution is a development of initial data on the singularity can be interpreted as saying that it is AVTD, in accordance with the BKL proposal and [33]. Third, if there are initial data on the singularity in a certain symmetry class, then it is natural to conjecture that the behavior is generically quiescent in this symmetry class, and that generic solutions are developments of initial data on the singularity. Conversely, if there are no initial data on the singularity, it is natural to conjecture that the generic behavior is oscillatory. Fourth, in quiescent settings, initial data on the singularity can potentially be used to parameterize solutions. In such settings, they thus fill a double role: they both provide as detailed information concerning the asymptotics as one could wish and parameterize solutions. However, there are at least two caveats to keep in mind. First, the formulation of the limits in (1) introduces a foliation dependence. It is natural to interpret this dependence as a lack of covariance. On the other hand, in many situations there are foliations that are uniquely determined by the spacetime. Assume that (M, g) is a globally hyperbolic spacetime with closed spacelike Cauchy hypersurfaces; satisfies the timelike convergence condition (this holds for the Einstein-scalar field equations with a vanishing cosmological constant); and has a crushing singularity (in the sense that there is a foliation by a family of Cauchy hypersurfaces whose mean curvatures diverge to infinity uniformly). Then (M, g) has a unique CMC crushing foliation in a neighborhood of the singularity. One natural covariant relation between the data on the singularity and the development is then to require that (1) holds for the uniquely determined CMC foliation. The notion of a cosmological time function, see [59], can be used similarly. Nevertheless, it is of interest to ask if the existence of one type of foliation inducing initial data $\mathfrak{I}$ on the singularity leads to the existence of other types of foliations inducing the same data $\mathfrak{I}$.

Second, even though a certain combination of matter model and symmetry class admits initial data on the singularity, it is unlikely that all solutions are developments of initial data on the singularity. In particular, a generalization of the notion of spikes appearing in the $\mathbb{T}^3$-Gowdy symmetric setting, see Section 3.2.1, can be expected to be relevant in the spatially inhomogeneous setting. Nevertheless, one can hope that these exceptional features only occur along hypersurfaces on the singularity; see [57] for an example.

3. Solutions with symmetry

One approach to the study of cosmological solutions to Einstein's equations is to start by studying solutions with a high degree of symmetry and then to gradually relax the symmetry conditions. In view of the fact that the starting point in standard cosmology is the cosmological principle, this is a quite natural way to proceed. The disadvantage of studying classes of symmetric solutions is that they are automatically non-generic. On the other hand, within a symmetry class, it might be possible to obtain a global understanding of the dynamics and to discover unexpected phenomena. Most, though not all, results in the absence of symmetry are stability results. In this case, the advantage is that one obtains results for open sets of initial data. On the other hand, in order to obtain results, one normally has to prove that the perturbed solutions remain close to the background. This means that one is unlikely to discover new phenomena.

3.1. *Spatially homogeneous solutions*

For a solution to be spatially homogeneous, it has to arise from homogeneous initial data. This means that the isometry group of the initial data has to be transitive. If the initial manifold Σ is 3-dimensional, this means that the universal covering manifold of Σ is either a 3-dimensional Lie group, or it is $\mathbb{S}^2 \times \mathbb{R}$. In the latter case, which is referred to as *Kantowski–Sachs*, the isometry group of simply connected initial data is the isometry group of the standard Riemannian metric on $\mathbb{S}^2 \times \mathbb{R}$ (note that this isometry group does not have a 3-dimensional transitive subgroup). If G is a 3-dimensional Lie group and $\mathfrak{g}$ is the corresponding Lie algebra, let, for $X \in \mathfrak{g}$, $\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}$ be defined by $\mathrm{ad}_X(Y) = [X, Y]$. Define $\eta_G \in \mathfrak{g}^*$ by $\eta_G(X) := \mathrm{tr}\ \mathrm{ad}_X$. If $\eta_G = 0$, then the Lie group is said to be *unimodular* (or of *Bianchi class A* in the cosmology literature). If $\eta_G \neq 0$, the Lie group is said to be *non-unimodular* (or of *Bianchi class B* in the cosmology literature). *Bianchi class A/B initial data* are left invariant initial data on a Bianchi class A/B Lie group. *Bianchi class A/B solutions* are solutions arising from Bianchi class A/B initial data. In what follows, we do not discuss the Kantowski–Sachs setting but rather focus on the Bianchi solutions. Imposing spatial homogeneity typically means that the combined Einstein-matter equations reduce to a system of ODE's. However, in the case of kinetic theory (the Einstein–Vlasov and Einstein–Boltzmann equations, e.g.), the equations are still PDE's.

As far as formulations of the equations are concerned, an important step in obtaining a unified framework for discussing Bianchi spacetimes was taken in [60], see also [61]. Many subsequent formulations, such as [62, 63], used to prove several of the results mentioned below, are based on the ideas of [60].

Since the literature in the spatially homogeneous setting is vast, we can here only mention an extremely limited selection of the results. Nevertheless, we refer to [60, 61, 64, 65] and the references cited therein for a description of the state of the art prior to the mid 90's; here our main focus will be on results obtained thereafter.

3.1.1. *The direction of the singularity*

Once the occurrence of a big bang and the high degree of isotropy of the CMB was established, a fundamental question immediately presented itself:

> ... if the 3 °K background radiation was last scattered at a redshift $z = 7$, then the radiation coming to us from two directions in the sky separated by more than about 30° was last scattered by regions of plasma whose prior histories had no causal relationship. [...] Robertson–Walker models therefore give no insight into why the observed microwave radiation from widely different angles in the sky has very precisely (0.2%) the same temperature;

see [66]. In order to address this problem, Misner suggested considering Bianchi type IX spacetimes. These are special examples of Bianchi class A spacetimes arising from left invariant initial data on a Lie group with a Lie algebra given by that of SU(2). Misner hoped that such spacetimes would have a different causal structure, allowing the establishment of a thermal equilibrium prior to the surface of last scattering. Beyond having the maximal number of degrees of freedom in the spatially homogeneous setting, and beyond playing a central role in the BKL proposal, Misner's expectation was thus that the Bianchi type IX spacetimes should play a central role in resolving the problem with high degree of isotropy of the CMB. In addition, there are vacuum Bianchi type IX solutions (namely the locally rotationally symmetric ones) such that the maximal Cauchy development can be extended through Cauchy horizons. In fact, there are two inequivalent maximal extensions, indicating that the initial data do not determine the spacetime; see [67, 68] for the original articles and [69, Theorem 25.4, p. 262] for a textbook presentation. Proving that inextendibility holds for initial data that are not locally rotationally symmetric is thus a natural goal and would correspond to proving strong cosmic censorship in the corresponding symmetry class.

The oscillatory setting. In the BKL proposal, the vacuum Bianchi type VIII and IX solutions are models for the expected oscillatory behavior (Bianchi type VIII solutions arise from left invariant initial data on Lie groups with a Lie algebra equal to that of $\mathrm{Sl}(2,\mathbb{R})$). That oscillations are to be expected in the vacuum Bianchi type VIII and IX settings can also be seen directly from Definition 1; there are no vacuum Bianchi type VIII or IX initial data on the singularity in the sense of Definition 1, since the Lie algebras in these cases are such that the p_D occurring in Definition 1 have to satisfy $2p_A = 1 + p_A - p_B - p_C > 0$ for all distinct A, B and C, which is not consistent with the requirements that the sum of the p_D equals one and the sum of the p_D^2 equals one. In spite of extensive discussions in the physics literature, the first mathematical result concerning the singularity in Bianchi type IX spacetimes was obtained in [70] (to the best of our knowledge). In particular, [70] contains a proof of the fact that, with the exception of the known locally rotationally solutions that give rise to Cauchy horizons through which the spacetimes can be extended, all Bianchi type VIII and IX vacuum initial data give rise to developments such that the Kretschmann scalar is unbounded in the incomplete directions of causal geodesics. This verifies both the curvature blow up conjecture and the C^2-version of the strong cosmic censorship conjecture in this setting. Moreover, the expansion normalized eigenvalues of the Weingarten map do not converge. In this sense, the solutions exhibit the oscillatory behavior conjectured by BKL. On the other hand, the result does not resolve the issue of the causal structure. Moreover, it does not yield the conclusion that solutions asymptote to an orbit of the BKL map.

Next, consider Einstein's equations coupled to an orthogonal perfect fluid with a linear equation of state $p = (\gamma - 1)\rho$, where ρ denotes the *energy density*, p the *pressure* and γ is a constant (here assumed to satisfy $\gamma \in (2/3, 2)$); the orthogonality assumption corresponds to the condition that fluid vector field be perpendicular to the hypersurfaces of spatial homogeneity. In this setting, generic Bianchi type IX solutions converge to an attractor on which the dynamics are determined by the BKL map; see [71] as well as [72, 73] for an overview of the state of the art in 2009. This is an indication that the BKL map is a good model for the dynamics. Unfortunately, the methods of [71] do not seem to be well suited to proving a similar result for Bianchi type VIII solutions. Nevertheless, in [74], the author proves that generic Bianchi type VIII and IX vacuum solutions converge to an attractor. Moreover, he proves that there is a set of initial data of full measure such that the corresponding solutions have particle horizons. On the other hand, he conjectures that for a Baire generic set of initial data, the corresponding solutions do not have particle horizons. The issue of the causality is thus quite subtle.

Turning to the question of whether solutions asymptote to an orbit of the Kasner map, the first results were obtained in [75, 76]; see also [77]. In these results, one fixes an orbit and then proves

that there are solutions converging to the orbit. In fact, in the first two results, the authors prove that there is a co-dimension one Lipschitz submanifold of the state space such that initial data in the submanifold yield solutions converging to the given orbit. Similar results were later obtained in the magnetic Bianchi type VI_0 setting in [78]. These results are certainly very interesting. However, they do not imply that there is a set of initial data of positive measure such that the corresponding solutions asymptote to an orbit of the Kasner map. However, in [79], the authors prove that there is a subset of the Kasner circle, say S, of full measure such that for each orbit originating with a point in S, there is a co-dimension one Lipschitz submanifold of the state space such that initial data in the submanifold yield solutions converging to the given orbit. Moreover, they prove that the union of all these Lipschitz submanifolds yields a subset of the state space with positive measure. The argument is long and technical, but in the end, one concludes that the set of initial data such that the corresponding solutions asymptote to an orbit of the Kasner map has positive measure. Clearly, this is an important step forward. However, it is still unclear if a similar statement holds for a subset of the state space of full measure.

Even though much attention has been devoted to Bianchi types VIII and IX, Bianchi type $\mathrm{VI}_{-1/9}$ is an equally general class of spatially homogeneous spacetimes. In fact, it can be argued that it is a more relevant model for oscillatory behavior in lower degrees of symmetry. However, due to the difficulty of analyzing the asymptotics of solutions to the relevant systems of ODE's, so far there are, to the best of our knowledge, no results. See, however, [80] for a formulation of the equations and a heuristic discussion of the asymptotics.

Oscillatory behavior can arise for two reasons; either because there are enough degrees of freedom (as in the case of Bianchi types $\mathrm{VI}_{-1/9}$, VIII and IX) or because of the presence of appropriate matter models. For examples of the latter, the reader is referred to, e.g., [78, 81, 82].

The convergent (or quiescent) setting. The behavior in the direction of the singularity can be convergent/quiescent mainly for two reasons. One possibility is that the matter fields are such that solutions naturally become convergent; this happens, e.g., in the presence of a scalar field or a stiff fluid. Another possibility is that the geometry is such that the mechanism causing oscillations is absent; this corresponds to condition (4) in Definition 1. In the case of orthogonal perfect fluids and Bianchi class A types different from VIII and IX, the situation was quite well understood already in [64, 83]. On the other hand, [63, 84] gives a good understanding of Bianchi class B orthogonal perfect fluid solutions which are not of type $\mathrm{VI}_{-1/9}$. In particular, [64,70,71,84,85] yield the C^2-version of strong cosmic censorship for non-stiff Bianchi class A and B spacetimes different from $\mathrm{VI}_{-1/9}$. The stiff fluid case is addressed in [71, 86, 87]. It is reasonable to expect convergent/quiescent solutions to arise from initial data on the singularity; see [57] for a justification of this statement in the Bianchi class A orthogonal stiff fluid setting.

Recollapse. Under certain circumstances, spacetimes are expected to recollapse, meaning that there is a singularity both to the future and to the past. More specifically, if the spatial topology of the spacetime is $\mathbb{S}^3$ or $\mathbb{S}^2 \times \mathbb{S}^1$, then the universe is in [88] conjectured to recollapse. This expectation goes under the name of the *closed universe recollapse conjecture*. In [89], the authors prove the closed universe recollapse conjecture in the case of diagonal Bianchi type IX universes satisfying the dominant energy condition and with non-negative average principal pressures. See also [24] for a discussion of this conjecture in more general settings.

3.1.2. *The expanding direction*

Turning to the expanding direction, it is natural to distinguish between non-accelerated expansion and accelerated expansion.

Non-accelerated expansion. For simple matter models with non-negative pressure and satisfying the dominant energy condition, one expects recollapse in the case of Kantowski–Sachs and Bianchi type IX. What remains to be considered is thus Bianchi class A types different from IX as well as Bianchi class B types. The Bianchi class A vacuum setting is discussed in [90–92]. Note that [91, 92] contains a discussion of the relation between the asymptotics and the conjectures formulated in Section 2.2. Since the relevant manifolds in the Bianchi class A setting are Seifert fibred, one expects collapse of the fibres after appropriate expansion normalization, and this is what happens.

It is also of interest to study the future asymptotics of solutions in the presence of matter fields. There is a very extensive literature on this subject. We here give a few examples. A fundamental question in cosmology is if one can deduce isotropy of the universe from the isotropy of the CMB. There is a result of this nature, due to Ehlers, Geren and Sachs, and referred to as the EGS theorem, see [93]. However, the theorem states that if the universe has a perfectly isotropic CMB, then the universe is spatially homogeneous and isotropic. It would be more interesting to know that if the CMB is almost isotropic, then the universe is almost isotropic. Such a result would be referred to as an almost-EGS theorem. However, that such a result cannot hold follows from [94–96] in which the authors study Bianchi type VII_0 solutions; see also [97]. In the above discussions, we have focused on orthogonal perfect fluids. However, it is also of interest to consider tilted perfect fluids; see [98–105] for some examples of results in this setting. For results in the presence of kinetic matter, see, e.g., [82, 106–110].

Accelerated expansion. The fundamental paper [111] on the asymptotics of Bianchi solutions to Einstein's equations with a positive cosmological constant establishes that, with the exception of Bianchi type IX, solutions asymptote to de Sitter space. Concerning the matter, it is only necessary to assume that the stress energy tensor obey the strong and dominant energy conditions. In the case of Bianchi type IX, a similar conclusion holds, assuming the cosmological constant is initially large enough relative to the spatial scalar curvature. This would seem to establish the cosmic no-hair conjecture in the spatially homogeneous setting. However, such a conclusion presupposes that the solution exists globally to the future. In the case of a specific matter model, this has to be established; see, e.g., [112]. After the appearance of observational support for accelerated expansion in the late universe, interest in mechanisms yielding accelerated expansion increased. In the mathematical literature, some corresponding examples of studies are given in [113–115].

3.2. *Spacetimes with a* 2*-dimensional isometry group*

A natural next step is to consider spacetimes with a 2-dimensional group of isometries. In this setting, even the issue of global existence of solutions is non-trivial, and the corresponding problem is a natural starting point. Once global existence has been established, it is then natural to try to analyze the asymptotics, prove strong cosmic censorship in the corresponding symmetry class etc. In what follows, when we speak of strong cosmic censorship, we always have the C^2-version in mind.

3.2.1. *Global existence, strong cosmic censorship*

One of the first examples of global existence for arbitrary initial data (in a symmetry class of the type of interest here), is the one concerning Gowdy spacetimes obtained in [13]. However, it was followed by a large collection of results, proving global existence of geometrically natural foliations for several different matter fields; see, e.g., [116–127]. The symmetry classes studied are Gowdy symmetry, $\mathbb{T}^2$-symmetry and surface symmetry. In the first two cases, the basic symmetry

assumption is that there is a compact and connected 2-dimensional Lie group of isometries acting effectively on the set of initial data on a closed 3-manifold. Due to [128, 129], this means that the group has to be $\mathbb{T}^2$. Moreover, the initial hypersurface must be $\mathbb{T}^3$, $\mathbb{S}^2 \times \mathbb{S}^1$, $\mathbb{S}^3$ or one of the Lens spaces (which means that a covering space of it is $\mathbb{S}^3$, so that we ignore Lens spaces from now on); see [117]. The 2-torus action on the initial hypersurface gives rise to two Killing vector fields of the corresponding development, say X_i, $i = 1, 2$. Then $\mathcal{T}_i := \epsilon_{\alpha\beta\gamma\delta} X_1^\alpha X_2^\beta \nabla^\gamma X_i^\delta$ are the associated *twist constants*, where $\epsilon_{\alpha\beta\gamma\delta}$ are the components of the volume form and ∇ is the Levi-Civita connection of the development. In the vacuum setting, the twist constants are constants, but for general matter models they are functions. The Gowdy setting corresponds to demanding that the twist constants vanish (this happens automatically in the case of $\mathbb{S}^2 \times \mathbb{S}^1$ and $\mathbb{S}^3$ spatial topology); see [130] for the original article by Gowdy. The possible topologies (ignoring the Lens spaces) are thus $\mathbb{T}^3$, $\mathbb{S}^2 \times \mathbb{S}^1$ and $\mathbb{S}^3$ in the Gowdy setting. The general case that the twist constants do not vanish is referred to as the $\mathbb{T}^2$-symmetric setting. In this case, the relevant spatial topology is $\mathbb{T}^3$. One can take a slightly more general perspective, insisting on only local Killing vector fields; see [123, 131]. In the surface symmetric setting, the topology of the spacetime is $\mathbb{R} \times \mathbb{S}^1 \times \mathbb{F}$, where $\mathbb{F}$ is a closed 2-manifold. Letting $\mathbb{G}$ denote the universal cover of $\mathbb{F}$, the isometry group of one of the standard metrics on 2-dimensional Euclidean, spherical or hyperbolic space should act on the last factor of $\mathbb{R} \times \mathbb{S}^1 \times \mathbb{G}$ by isometries of the lifted spacetime metric. The surface symmetric setting is only interesting outside of vacuum. Even though it is of course interesting to know that there is a global constant mean curvature foliation, for example, the corresponding results should be understood as a first step to analyzing the asymptotics in general and proving strong cosmic censorship in particular.

Strong cosmic censorship in the Gowdy setting. In [132] (see also [33]), the authors prove strong cosmic censorship for polarized Gowdy spacetimes. The Gowdy spacetimes have a two dimensional group of isometries, but the additional requirement of polarization means that the essential equations that need to be solved are linear. This significantly simplifies the analysis. Nevertheless, there are infinite dimensional sets of initial data leading to maximal Cauchy developments that are extendible. Proving that there is an open and dense subset of initial data whose maximal Cauchy developments are inextendible is therefore non-trivial.

A proof of strong cosmic censorship in the general $\mathbb{T}^3$-Gowdy symmetric vacuum setting can be found in [133–137] (these results can also be used to deduce strong cosmic censorship in the polarized $\mathbb{T}^3$-symmetric Einstein–Maxwell setting, see [138]). In this case, the essential equations can be thought of as wave map equations with hyperbolic space as a target. In particular, the equations are non-linear, and this causes substantial complications and rich dynamics. In particular, spikes appear. These features were first discovered numerically, see below, and later understood analytically, see [139] and [140]. Moreover, they are expected to be important more generally. Nevertheless, it is possible to prove that there is an open and dense set of initial data such that the corresponding maximal Cauchy developments are inextendible. In fact, quite detailed information can be derived concerning the asymptotics; see also [57] for an interpretation of the asymptotics in the direction of the singularity in terms of data on the singularity. Note, however, that strong cosmic censorship has neither been demonstrated in the $\mathbb{S}^2 \times \mathbb{S}^1$ nor in the $\mathbb{S}^3$ vacuum Gowdy-settings.

Strong cosmic censorship in the $\mathbb{T}^2$- and surface symmetric settings. In the vacuum $\mathbb{T}^2$-symmetric setting there are no results concerning strong cosmic censorship. This is due to the expectation that the behavior in the direction of the singularity should be oscillatory and therefore much more complicated; see the discussion of numerical work below for details. However, in the expanding direction there are some partial results as well as numerical simulations, see

below. On the other hand, including matter sometimes improves the situation. In order to understand why, it is useful to recall that in the case of Bianchi type IX vacuum spacetimes, e.g., the locally rotationally symmetric solutions are such that the maximal Cauchy development can be extended in inequivalent ways. They are therefore counterexamples to determinism. Due to the existence of these examples, proving strong cosmic censorship in the vacuum Bianchi type IX setting is subtle. Considering, on the other hand, Bianchi type IX solutions with non-trivial orthogonal perfect fluid matter, it can easily be demonstrated that the Ricci tensor contracted with itself blows up in the direction of the singularity for all solutions. Proving strong cosmic censorship in this setting is therefore trivial. In this sense, including matter can simplify the situation. Similarly, the reason proving strong cosmic censorship in the $\mathbb{T}^3$-Gowdy symmetric vacuum setting is non-trivial is that there are counterexamples to extendibility in that setting as well. Results concerning strong cosmic censorship in the surface symmetric setting with collisionless matter can be found in [141], see also [142, 143]. See also [144] for a future inextendibility result which covers, e.g., the Einstein–Vlasov equations in the $\mathbb{T}^2$-symmetric setting, including vacuum. A proof of strong cosmic censorship in the $\mathbb{T}^2$-symmetric Einstein–Vlasov setting with non-trivial Vlasov matter can be found in [145]. Finally, [146] extends this result to the case of a positive cosmological constant. Note, however, that [146] actually also covers the vacuum setting, assuming the area element of the group action does not vanish in the limit (in the direction of the singularity).

One bounce in the $\mathbb{T}^3$-Gowdy symmetric vacuum setting. To the best of our knowledge, there are no results in the oscillatory and spatially inhomogeneous setting. However, Warren Li recently proved a result describing one BKL bounce in the $\mathbb{T}^3$-Gowdy symmetric vacuum setting; see [147].

3.2.2. *The expanding direction*

Beyond proving strong cosmic censorship, it is of interest to derive detailed asymptotics in the expanding direction; in particular, it is of interest to verify the conjectures formulated in Section 2.2. In the case of polarized $\mathbb{T}^3$-Gowdy symmetric vacuum spacetimes, such an analysis can be found in [148, 149]. In fact, the asymptotics derived in [148] are detailed enough that they uniquely determine the solution, see [149]. The reason the analysis is so complete in this setting is that the essential equation is linear in this case. Note also that the conclusions follow as a very special case of the theory developed in [150]. In the case of general vacuum $\mathbb{T}^3$-Gowdy spacetimes, the future asymptotics are analyzed in [133, 151] (see also [138] for corresponding conclusion in the polarized $\mathbb{T}^3$-symmetric Einstein–Maxwell setting). In the $\mathbb{T}^2$-symmetric vacuum setting, there are general results that do, however, not provide very detailed information, see [152]. In the polarized $\mathbb{T}^2$-symmetric vacuum setting, there are results providing more detailed information, assuming one starts out close enough to an asymptotic regime, see [153]. There are also numerical studies of this situation, see below.

Including matter, there are results in the case of the Einstein–Vlasov equations with hyperbolic symmetry, see [126, 154]. In the case of the Einstein–Vlasov equations with surface symmetry and a positive cosmological constant, there are results in [155, 156]. In [25], the authors derive detailed information concerning the future asymptotics in the $\mathbb{T}^3$-Gowdy symmetric Einstein–Vlasov setting. In particular, the authors prove the cosmic no-hair conjecture in this setting.

3.2.3. *Specifying data on the singularity*

There is an extensive literature on specifying data on the singularity for classes of spacetimes with a 2-dimensional group of isometries, see, e.g., [45–51]. It would be interesting to relate these results to Definition 1. This has been done in [57] in the case of $\mathbb{T}^3$-symmetric Gowdy spacetimes. However, it remains to be done for the other cases. Related results concerning isotropic singularities can be found in [157–160].

3.3. *The* U(1)*-symmetric setting*

Finally, it is of interest to discuss the U(1)-symmetric setting; see [161] for a general review.

3.3.1. *Existence results and future asymptotics*

For general ideas concerning the reduction of the equations in the U(1)-symmetric setting, the reader is referred to [162]. The question of local existence, using variables that can be expected to be useful to treat global issues, is discussed in [163, 164]. Global existence is then discussed in [165]. In [166] (see also [167]), the authors prove a small data future global existence result which includes a proof of future causal geodesic completeness; the existence of a CMC foliation covering the causal future of the initial hypersurface; and the conclusion that the geometry asymptotes to that of a model spacetime. This result treats the so-called polarized case. An extension to the non-polarized case is given in [168, 169].

3.3.2. *Data on the singularity*

In the direction of the singularity, it is in the generic vacuum setting natural to expect oscillations, see the below discussion of numerical work. However, in the polarized setting, it is natural to expect quiescent asymptotics. In [53], the authors construct real analytic solutions satisfying these expectations using Fuchsian techniques. See also [52].

4. Future global non-linear stability in the expanding direction

In the cosmological setting, spacelike Cauchy hypersurfaces with mean curvature bounded away from zero appear naturally (in the absence of an asymptotically flat regime, hypersurfaces with vanishing mean curvature are, for many matter models, expected to be surrounded by hypersurfaces with mean curvature bounded away from zero, and hypersurfaces with significant positive mean curvature on some parts and significant negative mean curvature on other parts are unnatural in cosmology for the reasons given in the introduction). For a hypersurface satisfying this condition, the volume form is increasing (or decreasing) everywhere on the hypersurface. Moreover, due to Hawking's singularity theorem, the corresponding spacetime is either past or future timelike geodesically incomplete (assuming appropriate matter content). For these reasons, it is, when studying stability, natural to consider future expanding and future causally geodesically complete spacetimes and past contracting and past causally geodesically incomplete spacetimes in cosmology. When we speak of stability in the expanding direction, we typically assume that the background solution is future causally geodesically complete and has a foliation of the causal future of a Cauchy hypersurface such that each leaf of the foliation has strictly positive mean curvature. It is natural to divide the results that have been obtained according to the rate of the expansion, in particular according to whether the expansion is accelerated or not.

4.1. *Non-accelerated expansion*

In the case of non-accelerated expansion, it is, in accordance with the ideas described in Section 2.2, to be expected that the most favorable situation is when the spatial geometry is close to hyperbolic. Accordingly, the first future global non-linear stability result is the stability of the Milne model in the vacuum setting, see [170]; the metric of the Milne model is $-dt^2 + t^2\gamma$ on $(0,\infty) \times \Sigma$, where (Σ, γ) is a closed hyperbolic 3-manifold. The argument is based on considering energies arising from the Bel-Robinson tensor. This result was later generalized to higher dimensions in [171] using rougher energy estimates, and to different matter models, see, e.g., [172–174]. It is also worth noting that while asymptotic stability is obtained in [170], only stability is obtained

in [171], since there in higher dimensions is a non-trivial moduli space of negative Einstein manifolds; this adds an additional complication to the problem.

4.2. *Accelerated expansion*

The first result on future global non-linear stability in the cosmological setting is due to Helmut Friedrich, see [175]. In [175], the author proves, among other things, the global non-linear stability of de Sitter space. The method of proof is to derive a set of conformal field equations with respect to which the global stability problem becomes a local problem. However, this method is based on conformal invariance of the equations. The stability result [175] is extended to the Einstein–Maxwell–Yang–Mills setting in [176]. Similar methods have been used to obtain, e.g., the results [177–179].

The methods developed in [175] apply to specific matter models and specific spacetime dimensions. However, future global non-linear stability is expected to hold more generally. In [180], future global non-linear stability is demonstrated in the Einstein-non-linear scalar field setting. The particular situation considered in [180] is that of a potential with a strictly positive non-degenerate lower bound. However, the methods apply in any spacetime dimension ≥ 4. Moreover, they are based on coordinates similar to those used by Choquet-Bruhat to prove local existence in [6]. However, modifications are necessary; the contracted Christoffel symbols are replaced by appropriately chosen so-called *gauge source functions* and additional correction terms are added to the equations. Using the resulting equations, stability can be demonstrated using a bootstrap argument. Moreover, due to the causal structure of solutions, it is possible to prove stability results that are future global in time given only local assumptions in space. Consequently, it is possible to prove future global non-linear stability results for all spatially locally homogeneous solutions that exhibit de Sitter like expansion at the same time. The results in [180] were later generalized in [181] to the case of an exponential potential using similar methods. After [180] appeared, related results using similar methods (with additional complications due to the matter fields) appeared in, e.g., [25, 182–187]. In most of these references, the authors prove future global non-linear stability of specific spatially locally homogeneous solutions. However, in [25] the authors prove future global non-linear stability of arbitrary $\mathbb{T}^3$-Gowdy symmetric solutions to the Einstein–Vlasov system with a positive cosmological constant. In other words, the background solutions in this case do not only exhibit substantial anisotropies initially, they also exhibit substantial spatial inhomogeneities. Next, [186] contains, beyond stability results, a proof of the fact that for any choice of closed manifold for the spatial topology and any given choice of currently preferred standard model, there is a solution to the Einstein–Vlasov equations with a positive cosmological constant and the prescribed spatial topology such that observers in the corresponding spacetime cannot distinguish between their universe and the prescribed standard model; see [186] for details. In other words, given the currently preferred standard models of the universe, it is not possible to determine the global shape of the universe on the basis of observations.

Later, different methods were developed to prove future global non-linear stability in the presence of a positive cosmological constant, see [188]. In [188], the arguments rely on a conformal rescaling of the metric, but are still based on a particular choice of wave gauge, in analogy with [180]. The advantage of the methods developed in [188] is that the proof becomes substantially shorter. Moreover, the methods apply to perfect fluids with a linear equation of state of the form $p = K\rho$, where $0 < K \leq 1/3$. In other words, as opposed to earlier results concerning perfect fluids, the methods immediately include the radiation case. The methods developed in [188] have then been applied in, e.g., [189–191].

5. Big bang formation in the absence of symmetries

Prior to the work of Rodnianski and Speck, results concerning big bang singularities in the absence of symmetries were restricted to the ones obtained by specifying initial data on the singularity, see Section 2.5. The first stability result was obtained in [192], based on a study of the linearized equations in [193]. In [192], the authors prove that the occurrence of big bang singularities in spatially flat and isotropic solutions to the Einstein-scalar field equations (and to the Einstein-stiff fluid equations) is a past globally non-linearly stable phenomenon. The gauge used in the proof is determined by setting the shift vector field to zero and demanding that the time coordinate t equal $1/\theta$, where θ is the mean curvature of the constant-t hypersurface (we here use the opposite sign convention to that of [192] concerning the mean curvature). However, the geometric objects are expressed with respect to coordinate vector fields. The result [192] was later extended to prove past (future) global non-linear stability of the big bang (big crunch) in the case of spatially homogeneous and isotropic solutions with spherical spatial geometry and scalar field matter, see [194]. In [195], stable big bang formation is obtained in the case of hyperbolic spatial geometry. Moreover, the authors in this case also prove future global non-linear stability. In [196], the authors extend the results of [192] to cover the case of moderate anisotropies. In particular, [196] yields stable big bang formation for solutions to Einstein's vacuum equations with $\mathbb{T}^D$ spatial topology, assuming $D \geq 38$. In [192, 194, 196], the arguments are based on using a CMC foliation. This is very useful since it synchronizes the singularity. On the other hand, the CMC condition leads to an elliptic equation for the lapse function. In this sense, the gauge is non-local. On the other hand, the causal structure is such that particle horizons appear. In other words, two causal curves going into the singularity typically asymptotically lose the ability to communicate. This means that, from a causal point of view, the behavior localizes. It would therefore be of interest to prove a past global non-linear stability result which localizes in the spatial directions. This is done in [197]. In this paper, the authors use a different gauge which makes it possible to localize the analysis. The foliation the authors use in this case is determined by the condition that the scalar field be constant along the leaves of the foliation. In particular, the result requires the presence of a scalar field and therefore does not apply in the vacuum setting. In [198], the authors generalize this result to the presence of a perfect fluid with a linear equation of state $p = c_s^2\rho$, where $1/n < c_s^2 < 1$ and n is the spatial dimension (note that for $n = 3$, the case $0 \leq c_s^2 \leq 1/n$ is expected to be unstable in the direction of the singularity due to [199]). From these results, it is clear that the borderline case of a radiation fluid is of particular importance. It is therefore of interest to note that [200] yields stable big bang formation of the FLRW solutions to the Einstein–Vlasov-scalar field equations (the result simultaneously covers the spatially flat, hyperbolic and spherical settings); note that Vlasov matter behaves as a radiation fluid in the direction of the singularity.

In spite of the interest of the results mentioned above, they do not cover the regime one would naturally expect from, e.g., Definition 1; i.e., that a sufficient condition should be that the eigenvalues, say ℓ_A, of the expansion normalized Weingarten map satisfy $1 + \ell_A - \ell_B - \ell_C > 0$ for $B \neq C$. However, this is remedied in [201]. In this article, the authors prove stable big bang formation for the full range of spatially homogeneous and spatially flat solutions to the Einstein-scalar field equations for which one expects stability (in the polarized $U(1)$-symmetric setting, the conditions can even be relaxed, in accordance with condition (4) in Definition 1). Again, the authors use a CMC gauge with vanishing shift vector field. However, in [201] the authors use a Fermi–Walker propagated frame as opposed to considering the components of the geometry with respect to coordinate vector fields. Moreover, the equations are for the components of the second fundamental form with respect to the Fermi–Walker propagated frame, the connection coefficients of the frame, the components of the frame (and co-frame) with respect to coordinate

vector fields, the scalar field and the lapse function. The unknowns and the equations are therefore quite different from the ones in [192]. The advantage of the formulation in [201] is that the variables are such that one can formulate (and improve) quite mild bootstrap assumptions. Since the bootstrap assumptions are not very strong, it is easier to improve them. On the other hand, the conclusion that the bootstrap assumptions are satisfied does not yield that much information. Nevertheless, the information is sufficient to conclude that the eigenvalues of the expansion normalized Weingarten map converge, that the expansion normalized quantities associated with the scalar field converge and that the curvature blows up.

All the results mentioned above are stability results. In other words, the starting point is a given background solution (or family of background solutions). The conclusions of the results are then that perturbing initial data corresponding to the relevant background solution yields solutions that also have big bang singularities. In analogy with Hakwing's theorem, it would be desirable to prove that there is a general condition, which does not refer to a specific background solution, ensuring big bang formation with curvature blow up. Such a result is obtained in [202]. The result is in the Einstein-non-linear scalar field setting and the assumptions involve: bounds on the potential; the expected bounds on the eigenvalues of the expansion normalized Weingarten map; and, for a fixed bound on suitable Sobolev norms on the expansion normalized quantities (introduced in connection with (1)), a corresponding lower bound on the mean curvature (the conditions are such that, in particular, they are satisfied for all solutions with a CMC foliation that induce data on the singularity). The conclusions include past causal geodesic incompleteness and curvature blow up; see [202] for details. Since the results of [202] hold in the case of the Einstein-non-linear scalar field equations, they can be combined with the future global non-linear stability results in this setting, described in Section 4.2, in order to conclude past and future global non-linear stability of a large class of spatially locally homogeneous solutions. In particular, [202] applies in all the Einstein-scalar field settings in which stable big bang results were previously obtained. Many of the methods used in [202] are inspired by those developed in [201]. However, there are also significant differences. In particular, in [201] there is a background solution to prove stability of, and in the main theorem of [202] there is not. One important step in [202] is therefore to construct a substitute for a background solution.

6. Linear equations on cosmological backgrounds

There is a growing body of work concerning linear systems of wave equations on cosmological backgrounds. There is of course a vast physics literature on this subject, but we here focus on mathematical results. Some very early work is contained [203]. Since the central equation in the case of polarized Gowdy is linear, [132, 148, 149] are, simultaneously, studies of linear equations on cosmological backgrounds are directly relevant for the behavior of solutions to Einstein's equations. In [204], the authors study scalar perturbations of the Einstein–Euler equations. Further studies, mainly of the Klein–Gordon equation on spatially homogeneous backgrounds, are contained in, e.g., [85, 205–214]. Most of the results concern the equations on background geometries with convergent asymptotics. However, it is worth noting that [211] contains results in the case that the background geometry has chaotic dynamics. In [150], a general theory for linear systems of wave equations on cosmological backgrounds is developed. However, the geometries have to be such that the resulting equations are separable. The corresponding results cover many, though not all, of the previously mentioned results. For example, geometries which are similar to those of Kasner spacetimes are included (and solutions to systems of wave equations on such backgrounds are analyzed, both as far as past and future asymptotics are concerned). The main goal in [150] is to derive optimal energy estimates. This is done both with and without derivative losses. Moreover, in some situations it is demonstrated that there

is a homeomorphism between the regular initial data and asymptotic data. In [215], the case of spatially inhomogeneous backgrounds is considered. In this case, a priori assumptions concerning the geometry in the direction of a crushing singularity are made. Systems of linear wave equations on the corresponding backgrounds are then studied. In particular, optimal energy estimates and expansions of the solutions in the direction of the singularity are derived. However, the assumptions are general enough that the limits of the coefficients of the equations need not be continuous. This means that one has to localize along the causal future of a causal curve going into the singularity in order to derive detailed asymptotics.

In [216], conformal scattering of the Maxwell-scalar field system on de Sitter space is studied. Some aspects of [150] are improved in [217] and then applied to Maxwell's equations on a Kasner background (in the direction of the singularity). As already mentioned, the linearized Einstein-scalar field equations are studied in [193]. A scattering theory for the linearized Einstein-scalar field equations on a generalized Kasner background is developed in [218]; see also [219] for a result on scattering for the wave equation on de Sitter backgrounds in even spacetime dimensions.

7. Numerical work

There is a substantial literature of numerical results concerning cosmological solutions to Einstein's equations. It is not realistic to do it justice here. However, we wish to give some references that can be used as a starting point for those who wish to acquaint themselves with the literature.

7.1. *Results in symmetric settings*

Some of the early numerical work was carried out in [220], concerning Bianchi type IX solutions, and [221], concerning solutions with a 1- or 2-dimensional symmetry group. In fact, the numerical studies of solutions with symmetries ran in parallel with the mathematical studies described in Section 3.2 and 3.3. Moreover, the mathematical studies benefited from the numerical studies and vice versa. In [222], the algorithms in the Bianchi IX setting are improved and in [223] the authors probe the BKL proposal with numerical techniques. More detailed studies of quiescent and oscillatory behavior for cosmological singularities can be found in, e.g., [224–229]. Some studies concerning the future asymptotics in the $\mathbb{T}^2$-symmetric setting are to be found in [230, 231].

The idea that spikes, appearing, e.g., in $\mathbb{T}^3$-symmetric Gowdy spacetimes, occur naturally in the direction of the singularity arose due to numerical studies (the importance of spikes was later confirmed in mathematical results; cf., e.g., [137, 139]); see [140, 225, 232–237] for some references concerning spikes.

7.2. *Results in the absence of symmetries*

By now there is also a substantial literature in the absence of symmetries. Some of the early examples can be found in [238–241]. See [242, 243] for two more recent examples.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Funding

The author would like to acknowledge the support of Vetenskapsrådet, the Swedish Research Council, dnr. 2022-03053.

Acknowledgments

This article was in part written while the author was enjoying the hospitality of the Erwin Schrödinger Institute in Vienna.

References

[1] H. Nussbaumer, L. Bieri, *Discovering the Expanding Universe*, Cambridge University Press, Cambridge, 2009.
[2] S. W. Hawking, "The occurrence of singularities in cosmology. I", *Proc. R. Soc. Lond. A* **294** (1966), p. 511-521.
[3] S. W. Hawking, "The occurrence of singularities in cosmology. II", *Proc. R. Soc. Lond. A* **295** (1966), p. 490-493.
[4] B. O'Neill, *Semi-Riemannian geometry*, Pure and Applied Mathematics, vol. 103, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1983, With applications to relativity, xiii+468 pages.
[5] A. A. Penzias, R. W. Wilson, "A measurement of excess antenna temperature at 4080 Mc/s", *Astrophys. J. Lett.* **142** (1965), p. 419-421.
[6] Y. Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), p. 141-225.
[7] Y. Choquet-Bruhat, R. Geroch, "Global aspects of the Cauchy problem in general relativity", *Commun. Math. Phys.* **14** (1969), p. 329-335, http://projecteuclid.org/euclid.cmp/1103841822.
[8] Y. Fourès-Bruhat, "Théorèmes d'existence en mécanique des fluides relativistes", *Bull. Soc. Math. France* **86** (1958), p. 155-175, http://www.numdam.org/item?id=BSMF_1958__86__155_0.
[9] Y. Choquet-Bruhat, "Theorem of uniqueness and local stability for Liouville–Einstein equations", *J. Math. Phys.* **11** (1970), p. 3238-3243.
[10] Y. Choquet-Bruhat, "Un théorème d'existence pour le système intégrodifférentiel d'Einstein–Liouville", *C. R. Acad. Sci. Paris Sér. A-B* **271** (1970), p. A625-A628.
[11] Y. Choquet-Bruhat, "Problème de Cauchy pour le système intégro différentiel d'Einstein–Liouville", *Ann. Inst. Fourier (Grenoble)* **21** (1971), no. 3, p. 181-201, http://www.numdam.org/item?id=AIF_1971__21_3_181_0.
[12] D. Bancel, Y. Choquet-Bruhat, "Existence, uniqueness, and local stability for the Einstein–Maxwell–Boltzman system", *Commun. Math. Phys.* **33** (1973), p. 83-96, http://projecteuclid.org/euclid.cmp/1103859246.
[13] V. Moncrief, D. M. Eardley, "The global existence problem and cosmic censorship in general relativity", *Gen. Relativ. Gravit.* **13** (1981), no. 9, p. 887-892.
[14] A. E. Fischer, V. Moncrief, "Hamiltonian reduction and perturbations of continuously self-similar $(n+1)$-dimensional Einstein vacuum spacetimes", *Class. Quantum Gravity* **19** (2002), no. 21, p. 5557-5589.
[15] A. E. Fischer, V. Moncrief, "Conformal volume collapse of 3-manifolds and the reduced Einstein flow", in *Geometry, Mechanics, and Dynamics*, Springer, New York, 2002, p. 463-522.
[16] A. E. Fischer, V. Moncrief, "The reduced Einstein equations and the conformal volume collapse of 3-manifolds", *Class. Quantum Gravity* **18** (2001), no. 21, p. 4493-4515.
[17] A. E. Fischer, V. Moncrief, "The phase portrait of the reduced Einstein equations", *Int. J. Mod. Phys. D* **10** (2001), p. 825-831, Selected papers for the Gravity Research Foundation Annual Essay Competition 2001.
[18] A. E. Fischer, V. Moncrief, "Hamiltonian reduction, the Einstein flow, and collapse of 3-manifolds", *Nucl. Phys. B Proc. Suppl.* **88** (2000), p. 83-102, Constrained dynamics and quantum gravity, 1999 (Villasimius).
[19] A. E. Fischer, V. Moncrief, "Hamiltonian reduction of Einstein's equations and the geometrization of three-manifolds", in *International Conference on Differential Equations, Vol. 1, 2 (Berlin, 1999)*, World Scientific Publishing, River Edge, NJ, 2000, p. 279-282.
[20] A. E. Fischer, V. Moncrief, "The reduced Hamiltonian of general relativity and the σ-constant of conformal geometry", in *Mathematical and Quantum Aspects of Relativity and Cosmology (Pythagoreon, 1998)*, Lecture Notes in Physics, vol. 537, Springer, Berlin, 2000, p. 70-101.
[21] A. E. Fischer, V. Moncrief, "The Einstein flow, the σ-constant and the geometrization of 3-manifolds", *Class. Quantum Gravity* **16** (1999), no. 11, p. L79-L87.
[22] A. E. Fischer, V. Moncrief, "Hamiltonian reduction of Einstein's equations of general relativity", *Nucl. Phys. B Proc. Suppl.* **57** (1997), p. 142-161, Constrained dynamics and quantum gravity 1996 (Santa Margherita Ligure).
[23] M. T. Anderson, "Asymptotic behaviour of future-complete cosmological spacetimes", *Class. Quantum Gravity* **21** (2004), p. S11-S27, A spacetime safari: essays in honour of Vincent Moncrief.

[24] M. T. Anderson, "On long-time evolution in general relativity and geometrization of 3-manifolds", *Commun. Math. Phys.* **222** (2001), no. 3, p. 533-567.
[25] H. Andréasson, H. Ringström, "Proof of the cosmic no-hair conjecture in the $\mathbb{T}^3$-Gowdy symmetric Einstein–Vlasov setting", *J. Eur. Math. Soc. (JEMS)* **18** (2016), no. 7, p. 1565-1650.
[26] E. M. Lifshitz, I. M. Khalatnikov, "Investigations in relativistic cosmology", *Adv. Phys.* **12** (1963), no. 46, p. 185-249.
[27] E. M. Lifshitz, I. M. Khalatnikov, "Problems of relativistic cosmology", *Sov. Phys. Uspekhi* **6** (1964), no. 4, p. 495-522.
[28] V. A. Belinskii, E. M. Lifshitz, I. M. Khalatnikov, "Oscillatory approach to the singular point in relativistic cosmology", *Uspehi fiziceskih nauk* **102** (1970), no. 11, p. 463-500.
[29] I. M. Khalatnikov, E. M. Lifshitz, "General cosmological solution of the gravitational equations with a singularity in time", *Phys. Rev. Lett.* **24** (1970), no. 2, p. 76-79.
[30] V. A. Belinskii, I. M. Khalatnikov, E. M. Lifshitz, "Oscillatory approach to a singular point in the relativistic cosmology", *Adv. Phys.* **19** (1970), no. 80, p. 525-573.
[31] V. A. Belinskii, I. M. Khalatnikov, E. M. Lifshitz, "A general solution of the Einstein equations with a time singularity", *Adv. Phys.* **31** (1982), no. 6, p. 639-667.
[32] D. Eardley, E. Liang, R. Sachs, "Velocity-dominated singularities in irrotational dust cosmologies", *J. Math. Phys.* **13** (1972), p. 99-106.
[33] J. Isenberg, V. Moncrief, "Asymptotic behavior of the gravitational field and the nature of singularities in Gowdy spacetimes", *Ann. Phys.* **199** (1990), no. 1, p. 84-122.
[34] T. Damour, M. Henneaux, H. Nicolai, "Cosmological billiards", *Class. Quantum Gravity* **20** (2003), no. 9, p. R145-R200.
[35] T. Damour, H. Nicolai, "Higher-order M-theory corrections and the Kac-Moody algebra E_{10}", *Class. Quantum Gravity* **22** (2005), no. 14, p. 2849-2879.
[36] J. M. Heinzle, C. Uggla, N. Röhr, "The cosmological billiard attractor", *Adv. Theor. Math. Phys.* **13** (2009), no. 2, p. 293-407, http://projecteuclid.org/euclid.atmp/1234881637.
[37] J. M. Heinzle, C. Uggla, W. C. Lim, "Spike oscillations", *Phys. Rev. D* **86** (2012), no. 10, article no. 104049.
[38] J. Lott, "Kasner-like regions near crushing singularities", *Class. Quantum Gravity* **38** (2021), no. 5, article no. 055005.
[39] J. Lott, "On the initial geometry of a vacuum cosmological spacetime", *Class. Quantum Gravity* **37** (2020), no. 8, article no. 085017.
[40] H. Ringström, "On the geometry of silent and anisotropic big bang singularities", preprint, 2024, https://arxiv.org/abs/2101.04955.
[41] L. Andersson, A. D. Rendall, "Quiescent cosmological singularities", *Commun. Math. Phys.* **218** (2001), no. 3, p. 479-511.
[42] J. D. Barrow, "Quiescent cosmology", *Nature* **272** (1978), p. 211-215.
[43] J. Demaret, M. Henneaux, P. Spindel, "Nonoscillatory behaviour in vacuum Kaluza–Klein cosmologies", *Phys. Lett. B* **164** (1985), no. 1-3, p. 27-30.
[44] T. Damour, M. Henneaux, A. D. Rendall, M. Weaver, "Kasner-like behaviour for subcritical Einstein-matter systems", *Ann. Henri Poincaré* **3** (2002), no. 6, p. 1049-1111.
[45] S. Kichenassamy, A. D. Rendall, "Analytic description of singularities in Gowdy spacetimes", *Class. Quantum Gravity* **15** (1998), no. 5, p. 1339-1355.
[46] J. Isenberg, S. Kichenassamy, "Asymptotic behavior in polarized T^2-symmetric vacuum space-times", *J. Math. Phys.* **40** (1999), no. 1, p. 340-352.
[47] A. D. Rendall, "Fuchsian analysis of singularities in Gowdy spacetimes beyond analyticity", *Class. Quantum Gravity* **17** (2000), no. 16, p. 3305-3316.
[48] K. Anguige, K. P. Tod, "Isotropic cosmological singularities. I. Polytropic perfect fluid spacetimes", *Ann. Phys.* **276** (1999), no. 2, p. 257-293.
[49] F. Ståhl, "Fuchsian analysis of $S^2 \times S^1$ and S^3 Gowdy spacetimes", *Class. Quantum Gravity* **19** (2002), no. 17, p. 4483-4504.
[50] E. Ames, F. Beyer, J. Isenberg, P. G. LeFloch, "Quasilinear hyperbolic Fuchsian systems and AVTD behavior in T^2-symmetric vacuum spacetimes", *Ann. Henri Poincaré* **14** (2013), no. 6, p. 1445-1523.
[51] E. Ames, F. Beyer, J. Isenberg, P. G. LeFloch, "A class of solutions to the Einstein equations with AVTD behavior in generalized wave gauges", *J. Geom. Phys.* **121** (2017), p. 42-71.
[52] J. Isenberg, V. Moncrief, "Asymptotic behaviour in polarized and half-polarized U(1) symmetric vacuum space-times", *Class. Quantum Gravity* **19** (2002), no. 21, p. 5361-5386.
[53] Y. Choquet-Bruhat, J. Isenberg, V. Moncrief, "Topologically general U(1) symmetric vacuum space-times with AVTD behavior", *Nuovo Cimento Soc. Ital. Fis. B* **119** (2004), no. 7-9, p. 625-638.
[54] P. Klinger, "A new class of asymptotically non-chaotic vacuum singularities", *Ann. Phys.* **363** (2015), p. 1-35.
[55] G. Fournodavlos, J. Luk, "Asymptotically Kasner-like singularities", *Am. J. Math.* **145** (2023), no. 4, p. 1183-1272.
[56] H. Ringström, "Initial data on big bang singularities", preprint, 2022, https://arxiv.org/abs/2202.04919.

[57] H. Ringström, "Initial data on big bang singularities in symmetric settings", *Pure Appl. Math. Q* **20** (2024), no. 4, p. 1505-1539.
[58] A. Franco-Grisales, "Developments of initial data on big bang singularities for the Einstein-nonlinear scalar field equations", preprint, 2024, https://arxiv.org/abs/2409.17065.
[59] L. Andersson, G. J. Galloway, R. Howard, "The cosmological time function", *Class. Quantum Gravity* **15** (1998), no. 2, p. 309-322.
[60] G. F. R. Ellis, M. A. H. MacCallum, "A class of homogeneous cosmological models", *Commun. Math. Phys.* **12** (1969), p. 108-141, http://projecteuclid.org/euclid.cmp/1103841345.
[61] A. Krasiński, C. G. Behr, E. Schücking, F. B. Estabrook, H. D. Wahlquist, G. F. R. Ellis, R. Jantzen, W. Kundt, "The Bianchi classification in the Schücking–Behr approach", *Gen. Relativ. Gravit.* **35** (2003), no. 3, p. 475-489.
[62] J. Wainwright, L. Hsu, "A dynamical systems approach to Bianchi cosmologies: orthogonal models of class *A*", *Class. Quantum Gravity* **6** (1989), no. 10, p. 1409-1431.
[63] C. G. Hewitt, J. Wainwright, "A dynamical systems approach to Bianchi cosmologies: orthogonal models of class B", *Class. Quantum Gravity* **10** (1993), no. 1, p. 99-124.
[64] J. Wainwright, G. F. R. Ellis (eds.), *Dynamical Systems in Cosmology*, Cambridge University Press, Cambridge, 1997, Papers from the workshop held in Cape Town, June 27–July 2, 1994, xiv+343 pages.
[65] O. I. Bogoyavlensky, *Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics*, Springer Series in Soviet Mathematics, Springer-Verlag, Berlin, 1985, Translated from the Russian by Dmitry Gokhman, ix+301 pages.
[66] C. W. Misner, "Mixmaster Universe", *Phys. Rev. Lett.* **22** (1969), p. 1071-1074.
[67] C. W. Misner, A. Taub, "A singularity-free empty universe", *Sov. Phys. JETP* **28** (1969), p. 122-133.
[68] P. T. Chruściel, J. Isenberg, "Nonisometric vacuum extensions of vacuum maximal globally hyperbolic spacetimes", *Phys. Rev. D (3)* **48** (1993), no. 4, p. 1616-1628.
[69] H. Ringström, *The Cauchy Problem in General Relativity*, ESI Lectures in Mathematics and Physics, European Mathematical Society (EMS), Zürich, 2009, xiv+294 pages.
[70] H. Ringström, "Curvature blow up in Bianchi VIII and IX vacuum spacetimes", *Class. Quantum Gravity* **17** (2000), no. 4, p. 713-731.
[71] H. Ringström, "The Bianchi IX attractor", *Ann. Henri Poincaré* **2** (2001), no. 3, p. 405-500.
[72] J. M. Heinzle, C. Uggla, "Mixmaster: fact and belief", *Class. Quantum Gravity* **26** (2009), no. 7, article no. 075016.
[73] J. M. Heinzle, C. Uggla, "A new proof of the Bianchi type IX attractor theorem", *Class. Quantum Gravity* **26** (2009), no. 7, article no. 075015.
[74] B. Brehm, "Bianchi VIII and IX vacuum cosmologies: Almost every solution forms particle horizons and converges to the Mixmaster attractor", preprint, 2016, https://arxiv.org/abs/1606.08058.
[75] F. Béguin, "Aperiodic oscillatory asymptotic behavior for some Bianchi spacetimes", *Class. Quantum Gravity* **27** (2010), no. 18, article no. 185005.
[76] S. Liebscher, J. Härterich, K. Webster, M. Georgi, "Ancient dynamics in Bianchi models: approach to periodic cycles", *Commun. Math. Phys.* **305** (2011), no. 1, p. 59-83.
[77] M. Reiterer, E. Trubowitz, "The BKL conjectures for spatially homogeneous spacetimes", preprint, 2010, https://arxiv.org/abs/1005.4908.
[78] S. Liebscher, A. D. Rendall, S. B. Tchapnda, "Oscillatory singularities in Bianchi models with magnetic fields", *Ann. Henri Poincaré* **14** (2013), no. 5, p. 1043-1075.
[79] F. Béguin, T. Dutilleul, "Chaotic dynamics of spatially homogeneous spacetimes", *Commun. Math. Phys.* **399** (2023), no. 2, p. 737-927.
[80] C. G. Hewitt, J. T. Horwood, J. Wainwright, "Asymptotic dynamics of the exceptional Bianchi cosmologies", *Class. Quantum Gravity* **20** (2003), no. 9, p. 1743-1756.
[81] M. Weaver, "Dynamics of magnetic Bianchi VI_0 cosmologies", *Class. Quantum Gravity* **17** (2000), no. 2, p. 421-434.
[82] S. Calogero, J. M. Heinzle, "Oscillations toward the singularity of locally rotationally symmetric Bianchi type IX cosmological models with Vlasov matter", *SIAM J. Appl. Dyn. Syst.* **9** (2010), no. 4, p. 1244-1262.
[83] A. D. Rendall, "Global dynamics of the mixmaster model", *Class. Quantum Gravity* **14** (1997), no. 8, p. 2341-2356.
[84] K. Radermacher, "Strong cosmic censorship in orthogonal Bianchi class B perfect fluids and vacuum models", *Ann. Henri Poincaré* **20** (2019), no. 3, p. 689-796.
[85] H. O. Groeniger, "On Bianchi type VI_0 spacetimes with orthogonal perfect fluid matter", *Ann. Henri Poincaré* **21** (2020), no. 9, p. 3069-3094.
[86] K. Radermacher, "Orthogonal Bianchi B stiff fluids close to the initial singularity", preprint, 2017, https://arxiv.org/abs/1712.02699.
[87] H. O. Groeniger, "Quiescence for the exceptional Bianchi cosmologies", preprint, 2023, https://arxiv.org/abs/2311.05522.
[88] J. D. Barrow, G. J. Galloway, F. J. Tipler, "The closed-universe recollapse conjecture", *Mon. Not. R. Astron. Soc.* **223** (1986), p. 835-844.

[89] X.-F. Lin, R. M. Wald, "Proof of the closed-universe-recollapse conjecture for diagonal Bianchi type-IX cosmologies", *Phys. Rev. D (3)* **40** (1989), no. 10, p. 3280-3286.

[90] H. Ringström, "The future asymptotics of Bianchi VIII vacuum solutions", *Class. Quantum Gravity* **18** (2001), no. 18, p. 3791-3823.

[91] H. Ringström, "Future asymptotic expansions of Bianchi VIII vacuum metrics", *Class. Quantum Gravity* **20** (2003), no. 11, p. 1943-1989.

[92] J. M. Heinzle, H. Ringström, "Future asymptotics of vacuum Bianchi type VI_0 solutions", *Class. Quantum Gravity* **26** (2009), no. 14, article no. 145001.

[93] J. Ehlers, P. Geren, R. K. Sachs, "Isotropic solutions of the Einstein–Liouville equations", *J. Math. Phys.* **9** (1968), no. 9, p. 1344-1349.

[94] J. Wainwright, M. J. Hancock, C. Uggla, "Asymptotic self-similarity breaking at late times in cosmology", *Class. Quantum Gravity* **16** (1999), no. 8, p. 2577-2598.

[95] U. S. Nilsson, C. Uggla, J. Wainwright, W. C. Lim, "An almost isotropic cosmic microwave temperature does not imply an almost isotropic universe", *Astrophys. J.* **522** (1999), no. 1, p. L1-L3.

[96] U. S. Nilsson, M. J. Hancock, J. Wainwright, "Non-tilted Bianchi VII_0 models—the radiation fluid", *Class. Quantum Gravity* **17** (2000), no. 16, p. 3119-3134.

[97] H. Lee, E. Nungesser, J. Stalker, "On almost Ehlers–Geren–Sachs theorems", *Class. Quantum Gravity* **39** (2022), no. 10, article no. 105006.

[98] J. D. Barrow, S. Hervik, "The future of tilted Bianchi universes", *Class. Quantum Gravity* **20** (2003), no. 13, p. 2841-2854.

[99] S. Hervik, "The asymptotic behaviour of tilted Bianchi type VI_0 universes", *Class. Quantum Gravity* **21** (2004), no. 9, p. 2301-2317.

[100] A. A. Coley, S. Hervik, "Bianchi cosmologies: a tale of two tilted fluids", *Class. Quantum Gravity* **21** (2004), no. 17, p. 4193-4207.

[101] S. Hervik, R. van den Hoogen, A. Coley, "Future asymptotic behaviour of tilted Bianchi models of type IV and VII_h", *Class. Quantum Gravity* **22** (2005), no. 3, p. 607-633.

[102] S. Hervik, R. J. van den Hoogen, W. C. Lim, A. A. Coley, "Late-time behaviour of the tilted Bianchi type VI_h models", *Class. Quantum Gravity* **24** (2007), no. 15, p. 3859-3895.

[103] S. Hervik, R. J. van den Hoogen, W. C. Lim, A. A. Coley, "Late-time behaviour of the tilted Bianchi type $\mathrm{VI}_{-1/9}$ models", *Class. Quantum Gravity* **25** (2008), no. 1, article no. 015002.

[104] A. A. Coley, S. Hervik, "A note on tilted Bianchi type VI_h models: the type III bifurcation", *Class. Quantum Gravity* **25** (2008), no. 19, article no. 198001.

[105] S. Hervik, W. C. Lim, P. Sandin, C. Uggla, "Future asymptotics of tilted Bianchi type II cosmologies", *Class. Quantum Gravity* **27** (2010), no. 18, article no. 185006.

[106] A. D. Rendall, C. Uggla, "Dynamics of spatially homogeneous locally rotationally symmetric solutions of the Einstein–Vlasov equations", *Class. Quantum Gravity* **17** (2000), no. 22, p. 4697-4713.

[107] E. Nungesser, "Isotropization of non-diagonal Bianchi I spacetimes with collisionless matter at late times assuming small data", *Class. Quantum Gravity* **27** (2010), no. 23, article no. 235025.

[108] E. Nungesser, "Future non-linear stability for reflection symmetric solutions of the Einstein–Vlasov system of Bianchi types II and VI_0", *Ann. Henri Poincaré* **14** (2013), no. 4, p. 967-999.

[109] H. Lee, "Asymptotic behaviour of the relativistic Boltzmann equation in the Robertson–Walker spacetime", *J. Differ. Equ.* **255** (2013), no. 11, p. 4267-4288.

[110] H. Lee, "Global solutions of the Vlasov–Poisson–Boltzmann system in a cosmological setting", *J. Math. Phys.* **54** (2013), no. 7, article no. 073302.

[111] R. M. Wald, "Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant", *Phys. Rev. D (3)* **28** (1983), no. 8, p. 2118-2120.

[112] H. Lee, "Asymptotic behaviour of the Einstein–Vlasov system with a positive cosmological constant", *Math. Proc. Camb. Philos. Soc.* **137** (2004), no. 2, p. 495-509.

[113] A. D. Rendall, "Accelerated cosmological expansion due to a scalar field whose potential has a positive lower bound", *Class. Quantum Gravity* **21** (2004), no. 9, p. 2445-2454.

[114] A. D. Rendall, "Intermediate inflation and the slow-roll approximation", *Class. Quantum Gravity* **22** (2005), no. 9, p. 1655-1666.

[115] A. D. Rendall, "Dynamics of *k*-essence", *Class. Quantum Gravity* **23** (2006), no. 5, p. 1557-1569.

[116] J. Isenberg, V. Moncrief, "The existence of constant mean curvature foliations of Gowdy 3-torus spacetimes", *Commun. Math. Phys.* **86** (1982), no. 4, p. 485-493, http://projecteuclid.org/euclid.cmp/1103921839.

[117] P. T. Chruściel, "On space-times with $\mathrm{U}(1) \times \mathrm{U}(1)$ symmetric compact Cauchy surfaces", *Ann. Phys.* **202** (1990), no. 1, p. 100-150.

[118] P. T. Chruściel, *On Uniqueness in the Large of Solutions of Einstein's Equations ("Strong Cosmic Censorship"),*

Proceedings of the Centre for Mathematics and its Applications, Australian National University, vol. 27, Australian National University, Centre for Mathematics and its Applications, Canberra, 1991, 130 pages.

[119] A. D. Rendall, "Crushing singularities in spacetimes with spherical, plane and hyperbolic symmetry", *Class. Quantum Gravity* **12** (1995), no. 6, p. 1517-1533.

[120] A. D. Rendall, "Constant mean curvature foliations in cosmological space-times", *Helv. Phys. Acta* **69** (1996), p. 490-500, Journées Relativistes 96, Part II (Ascona, 1996).

[121] B. K. Berger, P. T. Chruściel, J. Isenberg, V. Moncrief, "Global foliations of vacuum spacetimes with T^2 isometry", *Ann. Phys.* **260** (1997), no. 1, p. 117-148.

[122] A. D. Rendall, "Existence and non-existence results for global constant mean curvature foliations", *Nonlinear Anal.* **30** (1997), no. 6, p. 3589-3598.

[123] A. D. Rendall, "Existence of constant mean curvature foliations in spacetimes with two-dimensional local symmetry", *Commun. Math. Phys.* **189** (1997), no. 1, p. 145-164.

[124] H. Andréasson, "Global foliations of matter spacetimes with Gowdy symmetry", *Commun. Math. Phys.* **206** (1999), no. 2, p. 337-365.

[125] J. Isenberg, M. Weaver, "On the area of the symmetry orbits in T^2 symmetric spacetimes", *Class. Quantum Gravity* **20** (2003), no. 16, p. 3783-3796.

[126] H. Andréasson, G. Rein, A. D. Rendall, "On the Einstein–Vlasov system with hyperbolic symmetry", *Math. Proc. Camb. Philos. Soc.* **134** (2003), no. 3, p. 529-549.

[127] H. Andréasson, A. D. Rendall, M. Weaver, "Existence of CMC and constant areal time foliations in T^2 symmetric spacetimes with Vlasov matter", *Commun. Partial Differ. Equ.* **29** (2004), no. 1-2, p. 237-262.

[128] P. S. Mostert, "On a compact Lie group acting on a manifold", *Ann. Math. (2)* **65** (1957), p. 447-455.

[129] P. S. Mostert, "Errata, "On a compact Lie group acting on a manifold"", *Ann. Math. (2)* **66** (1957), p. 589.

[130] R. H. Gowdy, "Vacuum spacetimes with two-parameter spacelike isometry groups and compact invariant hypersurfaces: topologies and boundary conditions", *Ann. Phys.* **83** (1974), p. 203-241.

[131] M. Tanimoto, "Locally U(1) × U(1) symmetric cosmological models: topology and dynamics", *Class. Quantum Gravity* **18** (2001), no. 3, p. 479-507.

[132] P. T. Chruściel, J. Isenberg, V. Moncrief, "Strong cosmic censorship in polarised Gowdy spacetimes", *Class. Quantum Gravity* **7** (1990), no. 10, p. 1671-1680.

[133] H. Ringström, "On a wave map equation arising in general relativity", *Commun. Pure Appl. Math.* **57** (2004), no. 5, p. 657-703.

[134] H. Ringström, "On Gowdy vacuum spacetimes", *Math. Proc. Camb. Philos. Soc.* **136** (2004), no. 2, p. 485-512.

[135] H. Ringström, "Asymptotic expansions close to the singularity in Gowdy spacetimes", *Class. Quantum Gravity* **21** (2004), p. S305-S322, A spacetime safari: essays in honour of Vincent Moncrief.

[136] H. Ringström, "Existence of an asymptotic velocity and implications for the asymptotic behavior in the direction of the singularity in T^3-Gowdy", *Commun. Pure Appl. Math.* **59** (2006), no. 7, p. 977-1041.

[137] H. Ringström, "Strong cosmic censorship in T^3-Gowdy spacetimes", *Ann. Math. (2)* **170** (2009), no. 3, p. 1181-1240.

[138] E. Nungesser, A. D. Rendall, "Strong cosmic censorship for solutions of the Einstein–Maxwell field equations with polarized Gowdy symmetry", *Class. Quantum Gravity* **26** (2009), no. 10, article no. 105019.

[139] A. D. Rendall, M. Weaver, "Manufacture of Gowdy spacetimes with spikes", *Class. Quantum Gravity* **18** (2001), no. 15, p. 2959-2975.

[140] W. C. Lim, "New explicit spike solutions—non-local component of the generalized mixmaster attractor", *Class. Quantum Gravity* **25** (2008), no. 4, article no. 045014.

[141] M. Dafermos, A. D. Rendall, "Strong cosmic censorship for surface-symmetric cosmological spacetimes with collisionless matter", *Commun. Pure Appl. Math.* **69** (2016), no. 5, p. 815-908.

[142] G. Rein, "Cosmological solutions of the Vlasov–Einstein system with spherical, plane, and hyperbolic symmetry", *Math. Proc. Camb. Philos. Soc.* **119** (1996), no. 4, p. 739-762.

[143] D. Tegankong, A. D. Rendall, "On the nature of initial singularities for solutions of the Einstein–Vlasov-scalar field system with surface symmetry", *Math. Proc. Camb. Philos. Soc.* **141** (2006), no. 3, p. 547-562.

[144] M. Dafermos, A. D. Rendall, "Inextendibility of expanding cosmological models with symmetry", *Class. Quantum Gravity* **22** (2005), no. 23, p. L143-L147.

[145] M. Dafermos, A. D. Rendall, "Strong cosmic censorship for surface-symmetric cosmological spacetimes with collisionless matter", *Commun. Pure Appl. Math.* **69** (2016), no. 5, p. 815-908.

[146] J. Smulevici, "Strong cosmic censorship for T^2-symmetric spacetimes with cosmological constant and matter", *Ann. Henri Poincaré* **9** (2008), no. 8, p. 1425-1453.

[147] W. Li, "BKL bounces outside homogeneity: Gowdy symmetric spacetimes", preprint, 2024, https://arxiv.org/abs/2408.12427.

[148] T. Jurke, "On future asymptotics of polarized Gowdy $\mathbf{T}^3$-models", *Class. Quantum Gravity* **20** (2003), no. 1, p. 173-191.

[149] H. Ringström, "Data at the moment of infinite expansion for polarized Gowdy", *Class. Quantum Gravity* **22** (2005), no. 9, p. 1647-1653.

[150] H. Ringström, "Linear systems of wave equations on cosmological backgrounds with convergent asymptotics", *Astérisque* **420** (2020), p. xi + 510.

[151] H. Ringström, "On curvature decay in expanding cosmological models", *Commun. Math. Phys.* **264** (2006), no. 3, p. 613-630.

[152] H. Ringström, "Instability of spatially homogeneous solutions in the class of $\mathbb{T}^2$-symmetric solutions to Einstein's vacuum equations", *Commun. Math. Phys.* **334** (2015), no. 3, p. 1299-1375.

[153] P. LeFloch, J. Smulevici, "Future asymptotics and geodesic completeness of polarized T^2-symmetric spacetimes", *Anal. PDE* **9** (2016), no. 2, p. 363-395.

[154] G. Rein, "On future geodesic completeness for the Einstein–Vlasov system with hyperbolic symmetry", *Math. Proc. Camb. Philos. Soc.* **137** (2004), no. 1, p. 237-244.

[155] S. B. Tchapnda, A. D. Rendall, "Global existence and asymptotic behaviour in the future for the Einstein–Vlasov system with positive cosmological constant", *Class. Quantum Gravity* **20** (2003), no. 14, p. 3037-3049.

[156] S. B. N. Tchapnda, N. Noutchegueme, "The surface-symmetric Einstein–Vlasov system with cosmological constant", *Math. Proc. Camb. Philos. Soc.* **138** (2005), no. 3, p. 541-553.

[157] K. Anguige, K. P. Tod, "Isotropic cosmological singularities and the Einstein–Vlasov equations", *Nonlinear Anal.* **30** (1997), no. 6, p. 3599-3603.

[158] K. Anguige, K. P. Tod, "Isotropic cosmological singularities. II. The Einstein–Vlasov system", *Ann. Phys.* **276** (1999), no. 2, p. 294-320.

[159] K. Anguige, "Isotropic cosmological singularities. III. The Cauchy problem for the inhomogeneous conformal Einstein–Vlasov equations", *Ann. Phys.* **282** (2000), no. 2, p. 395-419.

[160] K. Anguige, "A class of plane symmetric perfect-fluid cosmologies with a Kasner-like singularity", *Class. Quantum Gravity* **17** (2000), no. 10, p. 2117-2128.

[161] V. Moncrief, "Reflections on the U(1) problem in general relativity", *J. Fixed Point Theory Appl.* **14** (2013), no. 2, p. 397-418.

[162] V. Moncrief, "Reduction of the Einstein–Maxwell and Einstein–Maxwell–Higgs equations for cosmological spacetimes with spacelike U(1) isometry groups", *Class. Quantum Gravity* **7** (1990), no. 3, p. 329-352.

[163] Y. Choquet-Bruhat, V. Moncrief, "An existence theorem for the reduced Einstein equation", *C. R. Acad. Sci. Paris Sér. I Math.* **319** (1994), no. 2, p. 153-159.

[164] Y. Choquet-Bruhat, V. Moncrief, "Existence theorem for solutions of Einstein's equations with 1 parameter spacelike isometry groups", in *Quantization, Nonlinear Partial Differential Equations, and Operator Algebra (Cambridge, MA, 1994)*, Proceedings of Symposia in Pure Mathematics, vol. 59, American Mathematical Society, Providence, RI, 1996, p. 67-80.

[165] Y. Choquet-Bruhat, V. Moncrief, "Future complete Einsteinian space times with U(1) isometry group", *C. R. Acad. Sci. Paris Sér. I Math.* **332** (2001), no. 2, p. 137-144.

[166] Y. Choquet-Bruhat, V. Moncrief, "Future global in time Einsteinian spacetimes with U(1) isometry group", *Ann. Henri Poincaré* **2** (2001), no. 6, p. 1007-1064.

[167] Y. Choquet-Bruhat, V. Moncrief, "Nonlinear stability of an expanding universe with the S^1 isometry group", in *Partial Differential Equations and Mathematical Physics (Tokyo, 2001)*, Progress in Nonlinear Differential Equations and Their Applications, vol. 52, Birkhäuser, Boston, MA, 2003, p. 57-71.

[168] Y. Choquet-Bruhat, "Future complete S^1 symmetric Einsteinian spacetimes, the unpolarized case", *C. R. Math. Acad. Sci. Paris* **337** (2003), no. 2, p. 129-136.

[169] Y. Choquet-Bruhat, "Future complete U(1) symmetric Einsteinian spacetimes, the unpolarized case", in *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, Birkhäuser, Basel, 2004.

[170] L. Andersson, V. Moncrief, "Future complete vacuum spacetimes", in *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, Birkhäuser, Basel, 2004, p. 299-330.

[171] L. Andersson, V. Moncrief, "Einstein spaces as attractors for the Einstein flow", *J. Differ. Geom.* **89** (2011), no. 1, p. 1-47, http://projecteuclid.org/euclid.jdg/1324476750.

[172] L. Andersson, D. Fajman, "Nonlinear stability of the Milne model with matter", *Commun. Math. Phys.* **378** (2020), no. 1, p. 261-298.

[173] D. Fajman, Z. Wyatt, "Attractors of the Einstein–Klein-Gordon system", *Commun. Partial Differ. Equ.* **46** (2021), no. 1, p. 1-30.

[174] J. Wang, "Future stability of the 1 + 3 Milne model for the Einstein–Klein-Gordon system", *Class. Quantum Gravity* **36** (2019), no. 22, article no. 225010.

[175] H. Friedrich, "On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure", *Commun. Math. Phys.* **107** (1986), no. 4, p. 587-609, http://projecteuclid.org/euclid.cmp/1104116232.

[176] H. Friedrich, "On the global existence and the asymptotic behavior of solutions to the Einstein–Maxwell–Yang–Mills equations", *J. Differ. Geom.* **34** (1991), no. 2, p. 275-345, http://projecteuclid.org/euclid.jdg/1214447211.
[177] H. Friedrich, "Smooth non-zero rest-mass evolution across time-like infinity", *Ann. Henri Poincaré* **16** (2015), no. 10, p. 2215-2238.
[178] H. Friedrich, "Sharp asymptotics for Einstein-λ-dust flows", *Commun. Math. Phys.* **350** (2017), no. 2, p. 803-844.
[179] C. Lübbe, J. A. Valiente Kroon, "A conformal approach for the analysis of the non-linear stability of radiation cosmologies", *Ann. Phys.* **328** (2013), p. 1-25.
[180] H. Ringström, "Future stability of the Einstein-non-linear scalar field system", *Invent. Math.* **173** (2008), no. 1, p. 123-208.
[181] H. Ringström, "Power law inflation", *Commun. Math. Phys.* **290** (2009), no. 1, p. 155-218.
[182] C. Svedberg, "Future stability of the Einstein–Maxwell-scalar field system", *Ann. Henri Poincaré* **12** (2011), no. 5, p. 849-917.
[183] J. Speck, "The nonlinear future stability of the FLRW family of solutions to the Euler-Einstein system with a positive cosmological constant", *Selecta Math. (N.S.)* **18** (2012), no. 3, p. 633-715.
[184] X. Luo, J. Isenberg, "Power law inflation with electromagnetism", *Ann. Phys.* **334** (2013), p. 420-454.
[185] I. Rodnianski, J. Speck, "The nonlinear future stability of the FLRW family of solutions to the irrotational Euler-Einstein system with a positive cosmological constant", *J. Eur. Math. Soc. (JEMS)* **15** (2013), no. 6, p. 2369-2462.
[186] H. Ringström, *On the Topology and Future Stability of the Universe*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2013, xiv+718 pages.
[187] M. Hadžić, J. Speck, "The global future stability of the FLRW solutions to the dust-Einstein system with a positive cosmological constant", *J. Hyperbolic Differ. Equ.* **12** (2015), no. 1, p. 87-188.
[188] T. A. Oliynyk, "Future stability of the FLRW fluid solutions in the presence of a positive cosmological constant", *Commun. Math. Phys.* **346** (2016), no. 1, p. 293-312.
[189] C. Liu, T. A. Oliynyk, "Cosmological Newtonian limits on large spacetime scales", *Commun. Math. Phys.* **364** (2018), no. 3, p. 1195-1304.
[190] C. Liu, T. A. Oliynyk, "Newtonian limits of isolated cosmological systems on long time scales", *Ann. Henri Poincaré* **19** (2018), no. 7, p. 2157-2243.
[191] C. Liu, C. Wei, "Future stability of the FLRW spacetime for a large class of perfect fluids", *Ann. Henri Poincaré* **22** (2021), no. 3, p. 715-770.
[192] I. Rodnianski, J. Speck, "Stable big bang formation in near-FLRW solutions to the Einstein-scalar field and Einstein-stiff fluid systems", *Selecta Math. (N.S.)* **24** (2018), no. 5, p. 4293-4459.
[193] I. Rodnianski, J. Speck, "A regime of linear stability for the Einstein-scalar field system with applications to nonlinear big bang formation", *Ann. Math. (2)* **187** (2018), no. 1, p. 65-156.
[194] J. Speck, "The maximal development of near-FLRW data for the Einstein-scalar field system with spatial topology $\mathbb{S}^3$ ", *Commun. Math. Phys.* **364** (2018), no. 3, p. 879-979.
[195] D. Fajman, L. Urban, "Cosmic Censorship near FLRW spacetimes with negative spatial curvature", preprint, 2023, https://arxiv.org/abs/2211.08052.
[196] I. Rodnianski, J. Speck, "On the nature of Hawking's incompleteness for the Einstein-vacuum equations: the regime of moderately spatially anisotropic initial data", *J. Eur. Math. Soc. (JEMS)* **24** (2022), no. 1, p. 167-263.
[197] F. Beyer, T. A. Oliynyk, "Localized big bang stability for the Einstein-scalar field equations", *Arch. Ration. Mech. Anal.* **248** (2024), no. 1, article no. 3.
[198] F. Beyer, T. A. Oliynyk, "Past stability of FLRW solutions to the Einstein-Euler-scalar field equations and their big bang singularites", preprint, 2023, https://arxiv.org/abs/2308.07475.
[199] F. Beyer, E. Marshall, T. A. Oliynyk, "Past instability of FLRW solutions of the Einstein–Euler-scalar field equations for linear equations of state $p = K\rho$ with $0 \leq K < 1/3$", *Phys. Rev. D* **110** (2024), no. 4, article no. 044060.
[200] D. Fajman, L. Urban, "On the past maximal development of near-FLRW data for the Einstein scalar-field Vlasov system", preprint, 2024, https://arxiv.org/abs/2402.08544.
[201] G. Fournodavlos, I. Rodnianski, J. Speck, "Stable big bang formation for Einstein's equations: the complete sub-critical regime", *J. Amer. Math. Soc.* **36** (2023), no. 3, p. 827-916.
[202] H. O. Groeniger, O. Petersen, H. Ringström, "Formation of quiescent big bang singularities", preprint, 2023, https://arxiv.org/abs/2309.11370.
[203] S. Klainerman, P. Sarnak, "Explicit solutions of cmu = 0 on the Friedmann–Robertson–Walker space-times", *Ann. Inst. Henri Poincaré Sect. A (N.S.)* **35** (1981), no. 4, p. 253-257.
[204] P. T. Allen, A. D. Rendall, "Asymptotics of linearized cosmological perturbations", *J. Hyperbolic Differ. Equ.* **7** (2010), no. 2, p. 255-277.
[205] A. Galstian, T. Kinoshita, K. Yagdjian, "A note on wave equation in Einstein and de Sitter space-time", *J. Math. Phys.* **51** (2010), no. 5, article no. 052501.
[206] A. Vasy, "The wave equation on asymptotically de Sitter-like spaces", *Adv. Math.* **223** (2010), no. 1, p. 49-97.

[207] O. L. Petersen, "The mode solution of the wave equation in Kasner spacetimes and redshift", *Math. Phys. Anal. Geom.* **19** (2016), no. 4, article no. 26.
[208] A. Alho, G. Fournodavlos, A. T. Franzen, "The wave equation near flat Friedmann–Lemaître–Robertson–Walker and Kasner Big Bang singularities", *J. Hyperbolic Differ. Equ.* **16** (2019), no. 2, p. 379-400.
[209] A. Bachelot-Motet, A. Bachelot, "Waves on accelerating dodecahedral universes", *Class. Quantum Gravity* **34** (2017), no. 5, article no. 055010.
[210] P. M. Girão, J. Natário, J. D. Silva, "Solutions of the wave equation bounded at the big bang", *Class. Quantum Gravity* **36** (2019), no. 7, article no. 075016.
[211] H. Ringström, "A unified approach to the Klein-Gordon equation on Bianchi backgrounds", *Commun. Math. Phys.* **372** (2019), no. 2, p. 599-656.
[212] A. Bachelot, "Wave asymptotics at a cosmological time-singularity: classical and quantum scalar fields", *Commun. Math. Phys.* **369** (2019), no. 3, p. 973-1020.
[213] A. Bachelot, "Propagation of massive scalar fields in pre-big bang cosmologies", *Commun. Math. Phys.* **380** (2020), no. 2, p. 973-1001.
[214] D. Fajman, L. Urban, "Blow-up of waves on singular spacetimes with generic spatial metrics", *Lett. Math. Phys.* **112** (2022), no. 2, article no. 42.
[215] H. Ringström, "Wave equations on silent big bang backgrounds", preprint, 2021, https://arxiv.org/abs/2101.04939.
[216] G. Taujanskas, "Conformal scattering of the Maxwell-scalar field system on de Sitter space", *J. Hyperbolic Differ. Equ.* **16** (2019), no. 4, p. 743-791.
[217] A. F. Grisales, "Asymptotics of solutions to silent wave equations", preprint, 2023, https://arxiv.org/abs/2310.03582.
[218] W. Li, "Scattering towards the singularity for the wave equation and the linearized Einstein-scalar field system in Kasner spacetimes", preprint, 2024, https://arxiv.org/abs/2401.08437.
[219] S. Cicortas, "Scattering for the wave equation on de Sitter space in all even spatial dimensions", preprint, 2024, https://arxiv.org/abs/2309.07342.
[220] B. K. Berger, "Numerical study of initially expanding mixmaster universes", *Class. Quantum Gravity* **7** (1990), no. 2, p. 203-216.
[221] B. K. Berger, V. Moncrief, "Numerical investigation of cosmological singularities", *Phys. Rev. D (3)* **48** (1993), no. 10, p. 4676-4687.
[222] B. K. Berger, D. Garfinkle, E. Strasser, "New algorithm for Mixmaster dynamics", *Class. Quantum Gravity* **14** (1997), no. 2, p. L29-L36.
[223] B. K. Berger, D. Garfinkle, J. Isenberg, V. Moncrief, M. Weaver, "The singularity in generic gravitational collapse is spacelike, local and oscillatory", *Modern Phys. Lett. A* **13** (1998), no. 19, p. 1565-1573.
[224] B. K. Berger, V. Moncrief, "Numerical evidence that the singularity in polarized U(1) symmetric cosmologies on $T^3 \times \mathbf{R}$ is velocity dominated", *Phys. Rev. D (3)* **57** (1998), no. 12, p. 7235-7240.
[225] B. K. Berger, D. Garfinkle, "Phenomenology of the Gowdy universe on $T^3 \times \mathbf{R}$ ", *Phys. Rev. D (3)* **57** (1998), no. 8, p. 4767-4777.
[226] B. K. Berger, V. Moncrief, "Exact U(1) symmetric cosmologies with local mixmaster dynamics", *Phys. Rev. D (3)* **62** (2000), no. 2, article no. 023509.
[227] B. K. Berger, V. Moncrief, "Signature for local mixmaster dynamics in U(1) symmetric cosmologies", *Phys. Rev. D (3)* **62** (2000), no. 12, article no. 123501.
[228] B. K. Berger, J. Isenberg, M. Weaver, "Oscillatory approach to the singularity in vacuum spacetimes with T^2 isometry", *Phys. Rev. D (3)* **64** (2001), no. 8, article no. 084006.
[229] B. K. Berger, "Hunting local mixmaster dynamics in spatially inhomogeneous cosmologies", *Class. Quantum Gravity* **21** (2004), p. S81-S95, A spacetime safari: essays in honour of Vincent Moncrief.
[230] B. K. Berger, J. Isenberg, A. Layne, "Numerical study of the expanding direction of T^2-symmetric spacetimes", *Phys. Rev. D* **108** (2023), no. 10, article no. 104015.
[231] B. K. Berger, J. Isenberg, A. Layne, "Stability within T^2-symmetric expanding spacetimes", *Ann. Henri Poincaré* **21** (2020), no. 3, p. 675-703.
[232] D. Garfinkle, M. Weaver, "High velocity spikes in Gowdy spacetimes", *Phys. Rev. D (3)* **67** (2003), no. 12, article no. 124009.
[233] E. Nungesser, W. C. Lim, "The electromagnetic spike solutions", *Class. Quantum Gravity* **30** (2013), no. 23, article no. 235020.
[234] A. A. Coley, W. C. Lim, "Spikes and matter inhomogeneities in massless scalar field models", *Class. Quantum Gravity* **33** (2016), no. 1, article no. 015009.
[235] A. A. Coley, D. Gregoris, W. C. Lim, "On the first G_1 stiff fluid spike solution in general relativity", *Class. Quantum Gravity* **33** (2016), no. 21, article no. 215010.
[236] D. Gregoris, W. C. Lim, A. A. Coley, "Stiff fluid spike solutions from Bianchi type V seed solutions", *Class. Quantum Gravity* **34** (2017), no. 23, article no. 235013.

[237] W. C. Lim, "Numerical confirmations of joint spike transitions in G_2 cosmologies", *Class. Quantum Gravity* **39** (2022), no. 6, article no. 065001.

[238] D. Garfinkle, "Simulations of generic singularities in harmonic coordinates", in *The Conformal Structure of Spacetime*, Lecture Notes in Physics, vol. 604, Springer, Berlin, 2002, p. 349-358.

[239] D. Garfinkle, "Numerical simulations of generic singularities", *Phys. Rev. Lett.* **93** (2004), no. 16, article no. 161101.

[240] D. Garfinkle, "The nature of gravitational singularities", *Int. J. Mod. Phys. D* **13** (2004), no. 10, p. 2261-2265.

[241] D. Garfinkle, "Numerical simulations of general gravitational singularities", *Class. Quantum Gravity* **24** (2007), no. 12, p. S295-S306.

[242] D. Garfinkle, F. Pretorius, "Spike behavior in the approach to spacetime singularities", *Phys. Rev. D* **102** (2020), no. 12, article no. 124067.

[243] D. Garfinkle, A. Ijjas, P. J. Steinhardt, "Initial conditions problem in cosmological inflation revisited", *Phys. Lett. B* **843** (2023), article no. 138028.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 455-476
https://doi.org/10.5802/crmeca.288

Review article / *Article de synthèse*

Burnett's conjecture in general relativity

La conjecture de Burnett en relativité générale

Arthur Touati[a]

[a] CNRS & IMB, Bordeaux University, France
E-mail: arthur.touati@math.u-bordeaux.fr

Abstract. We present the literature on Burnett's conjecture in general relativity, which relate weak limits of vacuum solutions to relativistic kinetic theory. A special care is put on relating these works with early Choquet-Bruhat's results on high-frequency gravitational waves and geometric optics.

Résumé. Nous passons en revue la littérature sur la conjecture de Burnett en relativité générale, qui relie les limites faibles des solutions du vide à la théorie cinétique relativiste. Une attention particulière est portée sur le lien entre ces travaux et les premiers résultats de Choquet-Bruhat concernant les ondes gravitationnelles haute fréquence et l'optique géométrique.

Keywords. Einstein equations, Backreaction, Relativistic kinetic theory, Compensated compactness, Geometric optics, High-frequency gravitational waves.

Mots-clés. Équations d'Einstein, Rétroaction, Théorie cinétique relativiste, Compacité compensée, Optique géométrique, Ondes gravitationnelles haute fréquence.

Manuscript received 22 November 2024, revised 3 February 2025, accepted 6 February 2025.

1. Introduction

The author of the present text is more than honored to celebrate Yvonne Choquet-Bruhat's legacy in the field of mathematical general relativity, a field that, one could say, she launched in her seminal article of 1952 [1]. There, she built a local existence theory for the Einstein equations of general relativity. Of course, she did not stop there and went on to achieve an incredibly diverse career in general relativity, analysis and mathematical physics over more than 60 years. Her book [2] is the concrete proof of that, and fifteen years after its publication, this scientific testament still contains hidden treasures and research directions only waiting to be explored (see Section 6 for an example linked to this survey's topic). The goal of this survey is to present the mathematical literature on the so-called Burnett conjecture in general relativity. At this stage, we state the conjecture in the roughest way possible: weak limits of solutions to the Einstein vacuum equations can be identified with solutions to the massless Einstein–Vlasov system. Our first introductory task is to explicit the nature of the link between this conjecture and Choquet-Bruhat's work.

At first glance they seem rather unrelated, since the 785 pages of [2] don't contain any reference to Burnett himself, his conjecture or his original 1989 article [3]. However, and this may appease the reader rightfully questioning the presence of this text in this volume, Burnett's article [3] does contain a reference to Choquet-Bruhat's article [4] from 1969. In this article,

Choquet-Bruhat, following an approach pioneered by some physicists [5–8], has put on a solid mathematical footing an approximate geometric optics construction for these equations. Since an oscillating behaviour in the limit where the frequency tends to infinity is the simplest example of weak convergence in the mathematical sense, the approximate solutions of [4] fit into Burnett's abstract framework and provide a precursory and neat illustration of the so-called backreaction phenomenon, i.e. the effect of small scale inhomogeneities on a background. This connects Choquet-Bruhat's [4] and Burnett's [3] through the description of a common physics. But the link between these two works has recently been strengthened. Indeed, after being almost completely silent for decades, the mathematical activity around Burnett's conjecture has been revived by the first rigorous result obtained by Green and Wald in 2011 [9]. Since then, Choquet-Bruhat's early approximate construction from [4] has been revisited and its rigorous justification has proved to be an efficient way of attacking Burnett's conjecture.

Because of these two links, an early one from physics and a modern one from mathematics, the narrative of Burnett's conjecture will start in Section 2 by a description of [4]. Another, more prosaic, explanation behind this choice lies in the author's personal interests and his will to honor in this volume a relatively unknown, at least from the mathematical community, aspect of Choquet-Bruhat's work (corresponding to the eleventh chapter of [2]). We also refer the reader to her autobiography [10], where her work on high-frequency gravitational waves is mentioned several times.

Needless to say that the Burnett conjecture and its literature are also connected to many more areas of the mathematical theory of general relativity and of the partial differential equations' world. To highlight the richness of these connections is also one of the objectives of this text, which should be viewed as a complement to other reviews on Burnett's conjecture like the recent [11].

2. Choquet-Bruhat's article from 1969

As early as [12], physicists have tried to linearize the equations of general relativity around simple exact solutions of the Einstein vacuum equations, such as the Minkowski spacetime. This has led early general relativity physicists to think that spacetime itself could be a potential medium for a new kind of waves, called gravitational waves. We refer to Part VIII of [13] for a rich discussion of how gravitational waves propagate, how they are generated and how one can detect them. Chapter 35 of [13] is particularly relevant for the present discussion.

The obvious drawback of any linearization procedure is that it assumes that the perturbation is small, as well as all its derivatives, thus forbidding the description of more interesting physical situations where the gravitational field is strong. Moreover, the linearization procedure rules out the possibility of describing the "energy" of the gravitational waves themselves, since such a quantity, though complicated to define in general relativity, must be of quadratic nature. This has led physicists to imagine various averaging schemes (see for instance [5–8]) unfortunately lacking a rigorous mathematical framework. Thanks to her (at the time) recent experience in nonlinear geometric optics (see [14,15][1]), Yvonne Choquet-Bruhat was the first to rigorously applied the WKB expansion method to the Einstein vacuum equations themselves in the article [4]. We will discuss in more details the general strategy of geometric optics in Section 4.1.1, but for now we summarize the approach and main outcomes of this pioneer article.

In [4], Choquet-Bruhat considers a family of metrics of the form

$$\mathbf{g}_\lambda = \mathbf{g}_0 + \lambda \mathbf{g}^{(1)}\left(\frac{\varphi}{\lambda}\right) + \lambda^2 \mathbf{g}^{(2)}\left(\frac{\varphi}{\lambda}\right), \tag{1}$$

[1]This paper appears to be related to two other versions by the same or similar title, one earlier and one contemporaneous with the present entry. They are, respectively: [16] and [17].

where $\mathbf{g}_0$ is another metric, the phase φ is a scalar function and where by $\mathbf{g}^{(i)}\left(\varphi/\lambda\right)$ we actually mean $\mathbf{g}^{(i)}\left(x,\varphi(x)/\lambda\right)$. Note that we adapt her original notations in order to ease the comparison with more modern results on Burnett's conjecture (in particular those of Section 4.1). We assume moreover that each $\mathbf{g}^{(i)}$ is periodic with respect to the second variable. The metric $\mathbf{g}_\lambda$ should be thought as describing a vacuum spacetime in which a gravitational wave of wavelength λ is propagating. Crucially, the geometric optics expansion (1) allows the perturbation to be small but its derivatives to be large in the high-frequency limit $\lambda \to 0$. The goal is to find conditions on φ, $\mathbf{g}_0$, $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ such that $\mathbf{g}_\lambda$ is an approximate solution of the Einstein vacuum equations at order 1, that is $R_{\mu\nu}(\mathbf{g}_\lambda) = O(\lambda)$. For that, we compute the Ricci tensor of the metric $\mathbf{g}_\lambda$ defined by (1), and because of the curvature being second order with respect to the metric, it is of the form

$$R_{\mu\nu}(\mathbf{g}_\lambda) = \frac{1}{\lambda} R^{(-1)}_{\mu\nu} + R^{(0)}_{\mu\nu} + O(\lambda), \tag{2}$$

where $R^{(-1)}_{\mu\nu}$ depends on $\mathbf{g}_0$ and $\mathbf{g}^{(1)}$ and $R^{(0)}_{\mu\nu}$ depends on $\mathbf{g}_0$, $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$. In order to formulate the main result of [4], we introduce the polarization tensor of a tensor T with respect to a phase φ:

$$\mathbf{P}\left[T|\varphi\right]_\alpha := \mathbf{g}_0^{\mu\nu}\left(\partial_\mu\varphi T_{\nu\alpha} - \tfrac{1}{2}\partial_\alpha\varphi T_{\mu\nu}\right). \tag{3}$$

Theorem 1 ([4]). *Consider $\mathbf{g}_\lambda$ defined by (1) and $R^{(-1)}_{\mu\nu}$ and $R^{(0)}_{\mu\nu}$ as in (2).*

(i) $R^{(-1)}_{\mu\nu} = 0$ *if and only if* $\mathbf{g}_0^{-1}(\mathrm{d}\varphi,\mathrm{d}\varphi) = 0$ *and* $\mathbf{g}^{(1)}$ *satisfies the polarization condition*

$$\mathbf{P}\left[\mathbf{g}^{(1)}|\varphi\right] = 0. \tag{4}$$

(ii) *If* $R^{(-1)}_{\mu\nu} = 0$, *then* $R^{(0)}_{\mu\nu} = 0$ *if and only if* $\mathbf{g}_0$ *satisfies*

$$R_{\mu\nu}(\mathbf{g}_0) = \left\langle \tfrac{1}{4}\mathbf{g}_0^{\rho\sigma}\mathbf{g}_0^{\alpha\beta}\left(\partial_\theta\mathbf{g}^{(1)}_{\rho\alpha}\partial_\theta\mathbf{g}^{(1)}_{\sigma\beta} - \tfrac{1}{2}\partial_\theta\mathbf{g}^{(1)}_{\rho\sigma}\partial_\theta\mathbf{g}^{(1)}_{\alpha\beta}\right)\right\rangle_\theta \partial_\mu\varphi\partial_\nu\varphi, \tag{5}$$

$\mathbf{g}^{(1)}$ *satisfies the transport equation*

$$2\partial^\alpha\varphi\mathbf{D}_\alpha\mathbf{g}^{(1)} + (\Box_{\mathbf{g}}\varphi)\mathbf{g}^{(1)} = 0, \tag{6}$$

and $\mathbf{g}^{(2)}$ *satisfies a polarization condition of the form* $\mathbf{P}\left[\mathbf{g}^{(2)}|\varphi\right] = \mathbf{g}^{(1)}\star\mathbf{g}^{(1)}$, *where* $\star$ *denotes some tensor contractions.*

In (5), ∂_θ and $\langle\cdot\rangle_\theta$ denote respectively a derivative and the average with respect to the oscillating variable. Note that in Theorem 1, the case $\mathbf{g}_0^{-1}(\mathrm{d}\varphi,\mathrm{d}\varphi) \neq 0$ is excluded for physical reasons, since in this case a high-frequency change of variables can remove $\mathbf{g}^{(1)}$. Though Theorem 1 only produces approximate solutions to the Einstein vacuum equations, it provides a more rigorous averaging scheme than the results [5–8]. It shows how one can describe high-frequency gravitational waves in the nonlinear regime while still retaining some surprising linear features. For instance, as in the linearized gravity setting, the wave propagate along a null direction and the polarization condition 4 is actually equivalent to the TT-gauge conditions of linearized gravity which imply in particular that a gravitational wave has two degrees of freedom. Moreover, the (surprisingly linear) transport equation (6) is shown to propagate the polarization condition (4), so that the wave retains its form! Physically, this distinguishes greatly gravitational waves from other kind of waves, say in fluids, since it shows that the former don't self-steepen while the latter usually can accumulate and form shocks.

In addition to exhibiting this exceptional linear structure, Theorem 1 also shows the nonlinear effects caused by the wave on the background spacetime $\mathbf{g}_0$. Indeed, even though $\mathbf{g}_\lambda$ solves (approximately) the Einstein vacuum equations, this is not the case of $\mathbf{g}_0$ which "sees" the presence of the wave propagating, via the non-zero stress energy tensor in the right hand side of (5) (note that this expression, here derived rigorously, matches those found with other averaging schemes). In this sense, theorem 1 clearly prefigures Burnett's conjecture, which aims at characterizing completely the backreaction phenomenon (see Section 3 below).

We also mention subsequent works in the framework of [4], most of them being by Choquet-Bruhat herself. All these works show how important to her this direction of research was, from early to later stages of her career. High-frequency waves in relativistic fluids are considered in [14,15,18]. The coupling between gravitational waves and electromagnetic or fluid waves are the objects of [19] and [20] respectively. The case of a charged scalar field is considered in [21] and the one of the Yang–Mills equations in [22]. She even considers high-frequency gravitational waves with Gauss–Bonnet corrections or with stringy effects in [23] and [24] respectively. As we will see, the results on Burnett's conjecture discussed in Section 4.1 can also be thought as extensions of the pioneer work of Yvonne Choquet-Bruhat on high-frequency gravitational waves.

3. Statement of Burnett's conjecture

In this Section, we first introduce the massless Einstein–Vlasov system, state and discuss conceptualy Burnett's conjecture, and present the first rigorous result on the conjecture.

3.1. *Relativistic kinetic theory*

Relativistic kinetic theory is the name given to the physical description of a system of particles living in a spacetime described by general relativity, see the tenth chapter of [4] or [25]. While in classical kinetic theory the particles in consideration are most often gas or liquid particles, in general relativity, this perspective is adopted to describe "particles" at very different scales, such as stars, galaxies or even clusters of galaxies. Nevertheless, the framework of relativistic kinetic theory is identical to its classical counterpart: particles are described by a density function defined in phase space and solving a PDE encoding the interactions between the particles, and integrating this density with respect to the momentum variable provides macroscopic information on the system of particles.

If $(\mathscr{M}, \mathbf{g})$ is a Lorentzian manifold, we first look for an equivalent of the classical phase space encoding both position in space and time and velocity or momentum. Since tangent vectors to curves are represented by vector fields in differential geometry, the phase space in relativistic kinetic theory will be a particular subset of the tangent bundle $T\mathscr{M}$, the latter consisting of pairs (x, p) where $x \in \mathscr{M}$ and $p \in T_x\mathscr{M}$. More precisely, for $m \geq 0$ we define the mass shell at a point x by

$$\mathscr{P}_{m,x} := \left\{p \in T_x\mathscr{M} \mid \mathbf{g}\right\}_{|x}(p,p) = -m^2 \quad \text{and} \quad p \text{ is future oriented.}$$

If we set $\mathscr{P}_m := \bigsqcup_{x\in\mathscr{M}} \mathscr{P}_{m,x}$, then a density f describing particles of (rest) mass m is simply a function $f : \mathscr{P}_m \longrightarrow \mathbb{R}_+$, and $f(x,p)$ is the "number" of particles at position $x \in \mathscr{M}$ with momentum $p \in T_x\mathscr{M}$. To lighten the notations, in what follows we don't write down the measure on $\mathscr{P}_{m,x}$ induced by the metric $\mathbf{g}$.

We will assume here the absence of collisions between the particles in the system, so that the particles interact with each other only through their interaction with other fields. If moreover we neglect the charges of the particles, the only reasonable such field is the gravitational one and such a matter model is sometimes referred to as self-gravitating. This means that each particle is following a geodesic on $\mathscr{M}$, which is equivalent to the following Vlasov-type equation for the distribution function f:

$$p^\alpha \partial_\alpha f - \Gamma^\alpha_{\mu\nu} p^\mu p^\nu \partial_{p^\alpha} f = 0. \tag{7}$$

If $(\mathscr{M}, \mathbf{g})$ is given, the behaviour of the solutions to this equation is of tremendous physical importance. However, as always in general relativity, we can also describe the effect of a matter model on the Lorentzian metric $\mathbf{g}$ itself by plugging its stress energy tensor into the right hand

side of the Einstein equations. In the case of kinetic theory, the stress energy tensor is simply the second moment of the density f, which leads us to the so-called Einstein–Vlasov system with rest mass m:

$$\begin{cases} R_{\mu\nu}(\mathbf{g}) - \frac{1}{2}R(\mathbf{g})\mathbf{g}_{\mu\nu} = \int_{\mathscr{P}_{m,x}} f p_\mu p_\nu, \\ p^\alpha \partial_\alpha f - \Gamma^\alpha_{\mu\nu} p^\mu p^\nu \partial_{p^\alpha} f = 0. \end{cases} \tag{8}$$

In the context of (8), the Vlasov equation for f can also be interpreted as the condition for the stress energy tensor to have zero divergence. In the case $m = 0$, i.e. when the particles in consideration have zero rest mass, the system (8) is called the massless Einstein–Vlasov system. Note that in this case, the mass shell $\mathscr{P}_0$ is simply the collection of null cones for the metric $\mathbf{g}$ over each point of the manifold.

The literature on the system (8) is very rich and is nowadays concerned with global existence results, stationary solutions etc. Note that an important step was the proof of well-posedness in wave coordinates in [26]. Finally, as is standard in general relativity, we expect the limit $c \to +\infty$ to correspond to the classical Newtonian gravity. This is indeed true for relativistic kinetic theory, since it is proved in [27] that the famous Vlasov–Poisson system is the Newtonian limit of (8).

3.2. *First discussion of the conjecture*

We are now ready to state Burnett's conjecture from [3]. For that, fix a manifold $\mathscr{M}$ and consider sequences of metrics $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ solving the Einstein vacuum equations

$$R_{\mu\nu}(\mathbf{g}_\lambda) = 0, \tag{9}$$

and such that there exists another metric $\mathbf{g}_0$ such that in some coordinates system $(x^\alpha)_{\alpha=0,1,2,3}$ on $\mathscr{M}$ we have

$$(\mathbf{g}_\lambda)_{\alpha\beta} - (\mathbf{g}_0)_{\alpha\beta} \longrightarrow 0 \quad \text{uniformly}, \quad \partial_\mu\big((\mathbf{g}_\lambda)_{\alpha\beta} - (\mathbf{g}_0)_{\alpha\beta}\big) \longrightarrow 0 \quad \text{weakly}, \tag{10}$$

when the parameter λ tends to 0. Note that instead of the second assumption in (10), one can ask the sequence of first order derivatives $\big(\partial_\mu(\mathbf{g}_\lambda)_{\alpha\beta}\big)_{\lambda\in(0,1]}$ to be bounded.

Conjecture 2 (Burnett's conjecture). *One conjectures that:*

- *If $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ and $\mathbf{g}_0$ are as above, then $\mathbf{g}_0$ is a solution to the massless Einstein–Vlasov system (8) for a suitable density.*
- *Conversely, if $\mathbf{g}_0$ is a solution to the massless Einstein–Vlasov system (8), then there exists a sequence $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ as above and with limit $\mathbf{g}_0$.*

The above conjecture aims at completely describing backreaction for the Einstein vacuum equations, i.e. the effect of small scales inhomogeneities onto a background spacetime. These inhomogeneities are described by the lack of strong convergence of the first derivatives $\partial_\mu\big((\mathbf{g}_\lambda)_{\alpha\beta} - (\mathbf{g}_0)_{\alpha\beta}\big)$. In his original article, Burnett refers to this weak convergence as being the trace of some high-frequency oscillations in the metric itself. However, this regime also includes concentration effects, and only two references in the sequel of this text will allow for concentration effects, while all the others will be concerned with a true high-frequency regime, meaning where (10) is replaced by the stronger

$$\Big|\partial^k\big((\mathbf{g}_\lambda)_{\alpha\beta} - (\mathbf{g}_0)_{\alpha\beta}\big)\Big| \le \lambda^{1-k}, \quad k = 0, \dots, K,$$

for some $K \ge 2$.

The core idea behind Conjecture 2 is the following. If first order derivatives of $\mathbf{g}_\lambda$ converge only weakly, then products $\partial\mathbf{g}_\lambda\partial\mathbf{g}_\lambda$ of these derivatives don't necessarily converge to $\partial\mathbf{g}_0\partial\mathbf{g}_0$, and since

the Ricci tensor contains such semilinear terms, passing to the limit in (9) does not lead a priori to $R_{\mu\nu}(\mathbf{g}_0) = 0$ but rather to

$$R_{\mu\nu}(\mathbf{g}_0) - \tfrac{1}{2}R(\mathbf{g}_0)(\mathbf{g}_0)_{\mu\nu} =: T_{\mu\nu}^{\text{effective}},$$

where $T_{\mu\nu}^{\text{effective}}$ can be called the effective stress energy tensor of the inhomogeneities. The first part of Conjecture 2 can be thus rephrased as: there must exist some density function f defined on $\mathscr{P}_0$ (defined with the metric $\mathbf{g}_0$) such that

$$T_{\mu\nu}^{\text{effective}} = \int_{\mathscr{P}_{0,x}} f p_\mu p_\nu. \tag{11}$$

The effective stress energy tensor $T_{\mu\nu}^{\text{effective}}$ is thus caused by the lack of convergence of the semilinear terms $\partial\mathbf{g}_\lambda\partial\mathbf{g}_\lambda$. This shows that in the particular case (obviously included in Conjecture 2) where the convergence of $\partial_\mu\left((\mathbf{g}_\lambda)_{\alpha\beta} - (\mathbf{g}_0)_{\alpha\beta}\right)$ is actually strong, then we can pass to the limit in (9) and $\mathbf{g}_0$ thus solves the Einstein vacuum equations.

As explained in the introduction of [3], the fact that backreaction is necessarily of kinetic nature can be first observed on plane wave solutions to (9), where $\mathbf{g}_\lambda$ is explicitly given by

$$-\mathrm{d}u\mathrm{d}v + B_\lambda(u)^2\left(\mathrm{e}^{\lambda\alpha(u)\cos\left(\frac{u}{\lambda}\right)}\mathrm{d}x^2 + \mathrm{e}^{-\lambda\alpha(u)\cos(u/\lambda)}\mathrm{d}y^2\right),$$

where α is a given function of u and where B_λ needs to solve an ODE. One can show that $\mathbf{g}_\lambda$ converge in the sense (10) to $\mathbf{g}_0$ satisfying $R_{\mu\nu}(\mathbf{g}_0) = 1/2\alpha^2\partial_\mu u\partial_\nu u$. Formally summing such plane waves gives rise to the integrated expression in (11). However, and though Burnett does not mention it, the same kind of "discrete" effective stress energy tensor is obtained in the approximate geometric optics construction of [4] described in Section 2 (recall (5)). Burnett's explicit example has the benefit of being a family of exact solutions to (9), while Choquet-Bruhat's approximate solutions have the benefit of living outside any symmetry class. Moreover, as Section 4.1 will show, Choquet-Bruhat's strategy can be extended to produce exact solutions. However, Burnett's and Choquet-Bruhat's concrete examples of backreaction have one thing in common: the background kinetic spacetime metric $\mathbf{g}_0$ is a particular measure-valued solution of the massless Einstein–Vlasov system. More precisely, by considering several solutions of the eikonal equations $(u_{\mathbf{A}})_{\mathbf{A}\in\mathscr{A}}$ for $\mathscr{A}$ a finite set, we can consider the density function $f(x,p) = \sum_{\mathbf{A}\in\mathscr{A}} F_{\mathbf{A}}^2(x)\otimes\delta_{p=(\mathrm{d}u_{\mathbf{A}})^\#}$, where $\delta_{p=(\mathrm{d}u_{\mathbf{A}})^\#}$ denotes the Dirac measure at $(\mathrm{d}u_{\mathbf{A}})^\#$ in each tangent space and where the $F_{\mathbf{A}}$'s are scalar functions. In this case the massless Einstein–Vlasov becomes the so-called Einstein-null dusts system:

$$\begin{cases} R_{\mu\nu}(\mathbf{g}_0) = \displaystyle\sum_{\mathbf{A}\in\mathscr{A}} F_{\mathbf{A}}^2\partial_\mu u_{\mathbf{A}}\partial_\nu u_{\mathbf{A}}, \\ \mathbf{g}_0^{-1}(\mathrm{d}u_{\mathbf{A}},\mathrm{d}u_{\mathbf{A}}) = 0, \quad \text{for all } \mathbf{A}\in\mathscr{A}, \\ 2\mathbf{g}_0^{\alpha\beta}\partial_\beta u_{\mathbf{A}}\partial_\alpha F_{\mathbf{A}} + \left(\Box_{\mathbf{g}_0} u_{\mathbf{A}}\right)F_{\mathbf{A}} = 0, \quad \text{for all } \mathbf{A}\in\mathscr{A}, \end{cases} \tag{12}$$

The system (12) can be viewed as a "discrete" version of the massless Einstein–Vlasov system (8), but it also corresponds to the Einstein equations coupled to a perfect null fluid with zero pressure. Other examples of weak limits of vacuum spacetimes have been considered with two-dimensional symmetry, such as the $\mathbb{T}^2$ symmetry in [28].

3.3. *What does it say about the structure of the Einstein vacuum equations?*

Even before discussing actual results on Conjecture 2, we wish to reflect here on its significance. We start by noticing that, as it is apparent from its statement, Conjecture 2 is a double conjecture. Its two parts are of very different mathematical nature.

The first part, referred to as the *direct Burnett conjecture*, aims at describing the generic behaviour of sequences of solutions to the Einstein vacuum equations. In particular, it shows

that this set is not closed for the topology of (10). Since this backreaction is due to a lack of strong convergence, results on the direct Burnett conjecture fall into the scope of homogenization theory (we refer to [29] for an introduction to this field, which can be roughly defined as the study of averaging processes in disordered media), and the direct Burnett conjecture is an instance of compensated compactness. On the other hand, the second part of Conjecture 2, referred to as the *reverse Burnett conjecture*, aims at constructing solutions to the Einstein vacuum equations that converge in the sense of (10) to a given kinetic target. If it holds together with the direct part, then the set of massless kinetic spacetimes completely characterizes the closure of the set of solutions to the Einstein vacuum equations for the topology of (10). Therefore, if Burnett's conjecture is true, the Einstein vacuum equations stand in between two kinds of PDEs:

- PDEs for which the set of solutions is weakly closed, such as semilinear wave systems with the classical null condition of [30,31]. In this case, compensated compactness can be seen as a consequence of the famous div-curl lemma (see [32,33]). Such PDEs are said to display *rigidity*.
- PDEs for which weak limits of solutions are somehow arbitrary, such as incompressible fluid equations. In the case of incompressible Euler, weak limits of solutions are shown to solve the Euler–Reynolds system (see [34] or the survey [35]), which plays there a similar role as the massless Einstein–Vlasov. As opposed to the latter, the Euler–Reynolds system is underdetermined, thus leading to the display of *flexibility* (see the resolution of Onsager's conjecture in [36] for a consequence of this flexibility).

Therefore, one could say that the Einstein vacuum equations display some *flexible rigidity* or *rigid flexibility* (we let the reader decide which one is more accurate).

Not only does Burnett's conjecture distinguish the Einstein vacuum equations from incompressible fluid equations, in the sense that the failure of convergence satisfies a transport equation, it also reveals a hidden linear behaviour, despite the equations being very much nonlinear (already observed by Choquet-Bruhat in [4], see the discussion after Theorem 1). Indeed, the fact that backreaction can be described by the massless Einstein–Vlasov system, i.e. a self-gravitating matter model, shows that small scale inhomogeneities don't interact directly but only through their impact on the background metric. In other words, the linearity of the Vlasov equation with respect to the density, or the fact that no collision operator shows up in the right hand side, is a second feature of the Einstein vacuum equations. Depending on the context, these surprising nonlinear cancellations can be related to compensated compactness again (for instance in [11] it is described as a secondary form of compensated compactness), to the classical null condition and Shatah's normal form approach (see [37]) in semilinear waves or to transparency in geometric optics (see [38]). The deep link between all these approaches is discussed in depth in the Bourbaki seminar [39].

However, it is very well-known from the works on stability of Minkowski (see [40,41]) that the Einstein vacuum equations don't satisfy the null condition in wave coordinates. It is complicated to directly compare the structure of the Einstein vacuum equations in different gauges, but if Burnett's conjecture were to be true in full generality, it would show that they don't satisfy the null condition in any gauge or setting, since otherwise backreaction would not occur and vacuum spacetimes would not approach anything else but other vacuum spacetimes. Burnett's conjecture thus hints at a very fine structure: semilinear terms in the Einstein vacuum equations fail to satisfy the null condition (otherwise there would be no backreaction) but the failure of the null condition has a unique structure still leading to some cancellations (otherwise the Vlasov equation would be nonlinear).

In [42], Choquet-Bruhat introduced gauge-invariant definitions of the null condition and of a relaxation of it, but perhaps the most famous relaxed version of the null condition in

wave coordinates is the weak null condition introduced in [43]. One could then wonder if the aforementioned unique structure of the failure of the null condition is precisely Lindblad and Rodnianski's weak null condition. This is in fact not the case and if true, Burnett's conjecture thus demonstrates that the Einstein vacuum equations possess an even finer structure, strictly in between the null condition and the weak null condition. This last fact actually brings Burnett's conjecture closer to the bounded L^2 curvature theorem proved in [44], which (to the best of the author's knowledge) is the only other instance where the Einstein vacuum equations behave better than a system satisfying the weak null condition. The link between Burnett's conjecture and low-regularity solutions in general relativity is actually even stronger, but we postpone this discussion to Section 3.5.

3.4. *First rigorous result in the context of cosmology*

Before discussing in the rest of this survey the "PDE" results on Conjecture 2, we start with a result coming from the physics community. Obtained by Green and Wald in [9], this is the first mathematically rigorous result on Conjecture 2 since Burnett's original article [3].

Their motivation comes from cosmology, where a wealth of literature has been produced on the question of the effect of small scale inhomogeneities in the universe we live in. In particular, backreaction from these inhomogeneities onto the cosmological background has been believed to be a potential explanation of the presence of dark matter in our universe, in replacement of the standard explanation via the presence of (so far undetected) weakly-interacting "dark particles" or modifications of Einstein's gravity, see for instance [45,46]. However, if small scale inhomogeneities are described by sequences of metrics with the convergence regime (10), then Burnett's Conjecture 2 contradicts this belief, since the effect of backreaction is characterized by a massless Vlasov field. This motivated the work [9], where the authors get the following theorem:

Theorem 3 ([9]). *If $(\mathbf{g}_n)_{n\in\mathbb{N}}$ is a sequence of solutions to*

$$R_{\mu\nu}(\mathbf{g}_n) - \tfrac{1}{2}R(\mathbf{g}_n)(\mathbf{g}_n)_{\mu\nu} + \Lambda(\mathbf{g}_n)_{\mu\nu} = (T_n)_{\mu\nu},$$

where each T_n is trace free and satisfy the weak energy condition with respect to $\mathbf{g}_n$, and if $(\mathbf{g}_n)_{n\in\mathbb{N}}$ converges to some $\mathbf{g}_0$ as in (10), then

$$R_{\mu\nu}(\mathbf{g}_0) - \tfrac{1}{2}R(\mathbf{g}_0)(\mathbf{g}_0)_{\mu\nu} + \Lambda(\mathbf{g}_0)_{\mu\nu} = (T_0)_{\mu\nu},$$

where T_0 is trace free and satisfy the weak energy condition with respect to $\mathbf{g}_0$.

The conclusion of Theorem 3 completely rules out the possibility of dark matter being described by backreaction under the convergence regime (10) (see also [47] where the authors provide interesting explicit examples of such regime). Note also that the conclusion of Theorem 3 are in accordance with Burnett's conjecture, since the stress energy tensor of a massless Vlasov field is indeed traceless and satisfies the weak energy condition. Moreover, Theorem 3 fails to describe precisely the structure of the effective stress energy tensor, or the fact that the Vlasov transport equation is satisfied. However, some features distinguish it from the "PDE" results described in Sections 4 and 5 below: it allows for both cosmological constant and sequences of non-vacuum spacetimes, it does not require any symmetry nor gauge choice, it allows for both concentration and oscillation. For all these reasons, Theorem 3 reaches a high level of generality compared to the PDE results.

In the physics community, the article [9] has led to strong debates on the nature of backreaction, see for instance [48,49]. It also provoked an important increase in the number of citations of Burnett's original article [3], and it raised the interest of mathematicians and analysts working on the Einstein equations.

3.5. *Low-regularity and two different strategies*

In Sections 4 and 5 below, we will present the results [50–57] on Burnett's Conjecture 2. In order to motivate the splitting of the results between these two sections and their inner structures, we need to address a crucial feature of Burnett's conjecture, already hinted at in Section 3.3. To put it plainly, Burnett's conjecture is intrinsically a low-regularity phenomenon from the point of view of the Einstein vacuum equations. Indeed, the most generic well-posedness result for the Einstein vacuum equations is the celebrated bounded L^2 curvature theorem, which can be formulated as follows:

Theorem 4 ([44]). *The time of existence, with respect to a maximal foliation, of a classical solution to the Einstein vacuum equations depends only on the L^2 norm of the curvature and second fundamental form of the initial data set and on a lower bound of its volume radius.*

In other words, the Einstein vacuum equations are well-posed in H^s if $s \geq 2$. Theorem 4 is expected to be sharp, meaning that the *if* in the previous sentence is supposedly an *if and only if*[2]. However, in order to produce a nontrivial backreaction, a sequence $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ of solutions to the Einstein vacuum equations must necessarily be unbounded in any H^s for $s > 1$, since otherwise the convergence of $\partial_\mu\big((\mathbf{g}_\lambda)_{\alpha\beta} - (\mathbf{g}_0)_{\alpha\beta}\big)$ would be strong. This shows that Burnett's conjecture is a low-regularity phenomenon and that the mere existence of sequences $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ of solutions converging to some $\mathbf{g}_0$ on a time interval uniform in λ and being unbounded in H^s for $s > 1$ is not at all guaranteed by the general theory (said differently, the best we can do a priori is a time of existence of order λ). Since such sequences are the starting point of any theorems on the direct conjecture, they might be empty theorems. This leaves us with two possible strategies:

- First, if in a given setting (either a symmetry class or a particular gauge choice), no improvement of Theorem 4 is available, then one needs to construct concrete examples of such sequences unbounded in H^s for $s > 1$ bypassing the general obstruction, which is equivalent to proving the reverse conjecture! This motivates discussing first the reverse conjecture in the two frameworks fitting this category, and then the corresponding results on the direct conjecture (which are then ensured to be non-empty).
- Second, if in a given setting it is possible to somewhat improve Theorem 4, in the sense that the time of existence of a solution in this setting can be shown to depend only on the H^1 norm of the data in *some* directions, then the existence of the unbounded sequences would be ensured. As we will see, in this case proving the full Conjecture 2 actually reduces to proving its direct part.

Section 4 will present the results fitting into the first category, i.e. [50–54,56,57], in which Burnett's conjecture is studied in $\mathbb{U}(1)$ symmetry and generalized wave coordinates. Section 5 will present the only result fitting into the second catagory, i.e. [55], which is concerned with angularly regular spacetimes in double null gauge.

4. The $\mathbb{U}(1)$ symmetry and generalized wave coordinates

In this section, we discuss first the results [51,54,56,57], which attack the reverse conjecture via multiphase geometric optics constructions, and then the results [50,52,53], which prove the direct conjecture by means of microlocal defect measures.

[2]The presence of symmetries can lower the regularity threshold, see for instance Christodoulou's BV theory in spherical symmetry [58].

4.1. *The reverse conjecture*

As explained in Section 3.5, without any improvement of the general theory, proving the reverse conjecture amounts to the construction of particular solutions defined by special ansatz. As mentioned in Section 3.2, oscillation and concentration are the two different phenomena that, if present at the level of the first order derivatives of the metric, can cause boundedness in H^1 and unboundedness in H^s for $s > 1$ as required by the Burnett regime (10). If one chooses oscillation, then one actually studies the geometric optics approximation for the Einstein vacuum equations already considered by Choquet-Bruhat in [4].

4.1.1. *Geometric optics*

Broadly speaking, the field of geometric optics is concerned with the propagation of waves, and in particular with the description of rays out of it (see [59,60] for a more in depth introduction to this field). Concretely, the goal is to construct high-frequency solutions to nonlinear hyperbolic systems. These solutions are defined via multiscale asymptotic expansions of the form

$$u^\varepsilon \sim \varepsilon^p \sum_{n\geq 0} \varepsilon^n U_n\left(t, x, \frac{\varphi}{\varepsilon}\right), \tag{13}$$

where ε is a small parameter and where the U_n are periodic with respect to their third variable. This ansatz is very much in the spirit of WKB expansions (after Wentzel, Kramers and Brillouin) from the 1920's used to recover classical mechanics from quantum mechanics. In the case considered in [61,62] of a linear hyperbolic system, the oscillating profiles U_n can be shown to solve a hierarchy of transport equations along the "rays" defined by the phase φ, which needs to solve the eikonal equation associated to the differential operator. The first to have considered nonlinear systems is actually Yvonne Choquet-Bruhat herself, after having followed Leray's ("mon maitre", as she calls him in [10]) lectures at Princeton, see [14,15]. Of course, the article [4] is another instance of a geometric optics being applied to a nonlinear system, but with the subtlety that for the Einstein vacuum equations true hyperbolicity comes at a cost of a gauge choice. See also [63,64] for examples of geometric optics constructions on gauge-invariant semilinear systems. Given a nonlinear system, the basic questions are: can we construct formal solutions of the form (13)? are these approximate solutions stable, i.e. can we solve for a remainder in the expansion and transform them into exact solutions? In general, this is a very difficult problem strongly depending on the system and on the strength of the wave.

The parameter $p \geq 0$ in (13) describes the strength of the wave. If p is too large, no nonlinear effects will be observed in the transport hierarchy for the profiles. If one decreases the value of p, one can reach a particular value where nonlinear interactions terms enter the transport equation for the first profile, thus making it nonlinear. This is the regime of weakly nonlinear geometric optics, see [65,66] for classical results. For quadratic systems such as the Einstein equations, the critical value is $p = 1$. We remark crucially that [4] considers precisely the weakly nonlinear regime for the Einstein vacuum equations, and that if $p = 1$ then an ansatz of the form (13) precisely enters Burnett's convergence regime (10). In conclusion, we see that proving the reverse Burnett conjecture via oscillations is equivalent to proving the stability of the geometric optics approximation for the Einstein vacuum equations in the weakly nonlinear regime.

4.1.2. *The* $\mathbb{U}(1)$ *symmetry*

The first setting where the reverse Burnett conjecture has been studied is the one of $\mathbb{U}(1)$ symmetry, that is when the spacetime has a spacelike Killing vector field. The reduction of the 3+1 Einstein vacuum equations in this case can be deduced from the general Kaluza–Klein theory

(found for instance in Appendix VII of [2]). A general metric on a $3+1$ manifold with ∂_{x^3} Killing is of the form

$$\mathbf{g} = \mathrm{e}^{-2\varphi} g + \mathrm{e}^{2\varphi}\left(\mathrm{d}x^3 + A_\alpha \mathrm{d}x^\alpha\right),$$

where g is a $2+1$ Lorentzian metric, φ is a scalar function and A a 1-form. One can prove that the Einstein vacuum equations for $\mathbf{g}$ as above are equivalent to the following Einstein-wave map system in $2+1$ dimension

$$\begin{cases} \Box_g \varphi = -\frac{1}{2}\mathrm{e}^{-4\varphi} g^{-1}(\mathrm{d}\omega, \mathrm{d}\omega), \\ \Box_g \omega = 4 g^{-1}(\mathrm{d}\omega, \mathrm{d}\varphi), \\ R_{\mu\nu}(g) = 2\partial_\mu\varphi\partial_\nu\varphi + \frac{1}{2}\mathrm{e}^{-4\varphi}\partial_\mu\omega\partial_\nu\omega, \end{cases} \tag{14}$$

where ω is related to A. This system has been studied for instance in [67,68], motivated by its connection with cosmology, and the constraint equations for (14) are solved in [69]. Several choices of gauge are possible for (14), for instance generalized wave coordinates in [70] or null geodesic in [71], but it is the elliptic gauge which proves to be useful for Burnett's conjecture (or other low-regularity problems such as [72,73]). It is defined in the following way: after having decomposed g with lapse and shift, we ask its spacelike part to be conformal to the Euclidean metric (thanks to the uniformization theorem of surfaces) and each time slice to be maximal, i.e. with zero mean curvature. We now state the main result of [51] by Huneau and Luk:

Theorem 5 ([51]). *If $\left(g_0, \varphi_0, \omega_0, F_{\mathbf{A}}, u_{\mathbf{A}}\right)$, with $\mathbf{A}$ belonging to a finite set, is a generic solution on $[0,1]\times\mathbb{R}^2$ of the Einstein-null dusts system in $\mathbb{U}(1)$ symmetry in the elliptic gauge and close to Minkowski, then there exists a sequence $(g_\lambda, \varphi_\lambda, \omega_\lambda)_{\lambda\in(0,1]}$ of solutions of (14) in the elliptic gauge on $[0,1]\times\mathbb{R}^2$ and such that $(g_\lambda, \partial g_\lambda, \varphi_\lambda, \omega_\lambda) \longrightarrow (g_0, \partial g_0, \varphi_0, \omega_0)$ uniformly and $(\partial\varphi_\lambda, \partial\omega_\lambda) \longrightarrow (\partial\varphi_0, \partial\omega_0)$ weakly when λ tends to 0.*

The Einstein-null dusts system in $\mathbb{U}(1)$ symmetry mentioned in Theorem 5 is the equivalent of (14) with the addition of the stress energy tensor of the null dusts system, see (12). The existence of the background solution $\left(g_0, \varphi_0, \omega_0, F_{\mathbf{A}}, u_{\mathbf{A}}\right)$ follows from a low-regularity existence result in elliptic gauge from [74]. If we denote by $\mathfrak{g}$ either the lapse, the shift or the conformal factor, and by ψ either φ or ω, the system (14) in elliptic gauge can be rewritten in the following schematic form

$$\begin{cases} \Box_g \psi = g^{-1}(\mathrm{d}\psi, \mathrm{d}\psi), \\ \Delta\mathfrak{g} = (\partial\psi)^2 + (\nabla\mathfrak{g})^2. \end{cases} \tag{15}$$

The sequence $(\mathfrak{g}_\lambda, \psi_\lambda)_{\lambda\in(0,1]}$ is defined by a high-frequency ansatz, a simplified version of which being

$$\begin{aligned} \psi_\lambda &= \psi_0 + \lambda\sum_{\mathbf{A}} \psi_{\mathbf{A}}^{(1)}\left(\frac{u_{\mathbf{A}}}{\lambda}\right) + \lambda^2\left(\sum_{\mathbf{A}} \psi_{\mathbf{A}}^{(2)}\left(\frac{u_{\mathbf{A}}}{\lambda}\right) + \sum_{\mathbf{A}\neq\mathbf{B},\pm} \psi_{\mathbf{AB}}^{(2,\pm)}\left(\frac{u_{\mathbf{A}} \pm u_{\mathbf{B}}}{\lambda}\right)\right) + \text{remainder}, \\ \mathfrak{g}_\lambda &= \mathfrak{g}_0 + \lambda^2\left(\sum_{\mathbf{A}} \mathfrak{g}_{\mathbf{A}}^{(2)}\left(\frac{u_{\mathbf{A}}}{\lambda}\right) + \sum_{\mathbf{A}\neq\mathbf{B},\pm} \mathfrak{g}_{\mathbf{AB}}^{(2,\pm)}\left(\frac{u_{\mathbf{A}} \pm u_{\mathbf{B}}}{\lambda}\right)\right) + \text{remainder}, \end{aligned} \tag{16}$$

where, for instance, by $f(u_{\mathbf{A}}/\lambda)$ we mean that f is a quantity depending in a periodic manner on $u_{\mathbf{A}}/\lambda$ (as the profiles U_n in (13)). By plugging (16) in (15), we derive transport equations for the profiles $\psi^{(i)}$ and "algebraic" equations for the profiles $\mathfrak{g}^{(i)}$ (the solvability of the latter requires a coherence assumption on the $u_{\mathbf{A}}$'s, as it is common in multiphase geometric optics). Solving for the remainders in (16) over the required time scale is the most technical part of the proof and relies crucially on the structure of (15). From the present perspective, this structure can be summarized as follows: the dynamics and the geometry are completely decoupled into a quasilinear hyperbolic system with the null condition and a semilinear elliptic system without the null condition. This observation explains why there is no λ term in $\mathfrak{g}_\lambda$, or equivalently why $\partial\mathfrak{g}_\lambda$ converges strongly, as opposed to $\partial\psi_\lambda$. It also sheds light on the discussion of Section 3.3: the lack of null condition is compensated here by an elliptic operator, very much in the spirit of the

Coulomb gauge used in [44]. To conclude the discussion of Theorem 5, let us clarify how dust-like backreaction can arise. If say $\psi^{(1)}_{\mathbf{A}}(u_{\mathbf{A}}/\lambda) = \cos(u_{\mathbf{A}}/\lambda)\,\tilde{\psi}^{(1)}_{\mathbf{A}}$, then the quadratic term $\partial_\mu\psi_\lambda\partial_\nu\psi_\lambda$ in (14) or (15) becomes

$$\begin{aligned}\partial_\mu\psi_\lambda\partial_\nu\psi_\lambda \;=\; & \partial_\mu\psi_0\partial_\nu\psi_0 + \frac{1}{2}\sum_{\mathbf{A},\mathbf{B}}\left(\tilde{\psi}^{(1)}_{\mathbf{A}}\right)^2\partial_\mu u_{\mathbf{A}}\partial_\nu u_{\mathbf{A}} \\ & -\frac{1}{2}\sum_{\mathbf{A}}\sin\left(\frac{u_{\mathbf{A}}}{\lambda}\right)\tilde{\psi}^{(1)}_{\mathbf{A}}\partial_{(\mu}\psi_0\partial_{\nu)}u_{\mathbf{A}} - \frac{1}{2}\sum_{\mathbf{A},\mathbf{B}}\left(\tilde{\psi}^{(1)}_{\mathbf{A}}\right)^2\cos\left(\frac{2u_{\mathbf{A}}}{\lambda}\right)\partial_\mu u_{\mathbf{A}}\partial_\nu u_{\mathbf{A}} \\ & +\sum_{\mathbf{A}\neq\mathbf{B}}\sin\left(\frac{u_{\mathbf{A}}}{\lambda}\right)\sin\left(\frac{u_{\mathbf{B}}}{\lambda}\right)\tilde{\psi}^{(1)}_{\mathbf{A}}\tilde{\psi}^{(1)}_{\mathbf{B}}\partial_\mu u_{\mathbf{A}}\partial_\nu u_{\mathbf{B}} + O(\lambda)\,.\end{aligned}$$

Thanks to the coherence assumption, the last two lines converge weakly to 0 when λ tends to 0, showing how a new resonant term appears in the high-frequency limit.

In the forthcoming [54], Huneau and Luk prove that it is possible to rigorously send the number of dusts to infinity. Relying on an improved existence result in this setting [75], they perform a double limit argument, i.e. construct ansatz of the form (16) with uniform estimates with respect to both λ and the number of dusts. In particular, this allows to target general solutions to the massless Einstein–Vlasov system, by approximating any measure which is absolutely continuous with respect to the Lebesgue measure by sums of Dirac measures.

4.1.3. *Generalized wave coordinates*

In this section, we discuss results on the reverse Burnett conjecture for the Einstein vacuum equations in generalized wave coordinates. To define them, we give the expression of the Ricci tensor in any coordinates system:

$$2R_{\alpha\beta}(\mathbf{g}) = -\tilde{\Box}_{\mathbf{g}}\mathbf{g}_{\alpha\beta} + \mathbf{g}_{\rho(\alpha}\partial_{\beta)}H^\rho + H^\rho\partial_\rho\mathbf{g}_{\alpha\beta} + P_{\alpha\beta}(\mathbf{g})(\partial\mathbf{g},\partial\mathbf{g}), \tag{17}$$

where $\tilde{\Box}_{\mathbf{g}} := \mathbf{g}^{\mu\nu}\partial_\mu\partial_\nu$, $H^\rho := \mathbf{g}^{\mu\nu}\Gamma^\rho_{\mu\nu}$ and $P_{\alpha\beta}(\mathbf{g})(\partial\mathbf{g},\partial\mathbf{g})$ contains every quadratic semilinear terms. Constructing coordinates such that $H = 0$ while simultaneously solving the reduced equations $\tilde{\Box}_{\mathbf{g}}\mathbf{g}_{\alpha\beta} = P_{\alpha\beta}(\mathbf{g})(\partial\mathbf{g},\partial\mathbf{g})$ has been shown to be possible in the pioneering [1]. Since $H^\rho = -\Box_{\mathbf{g}}x^\rho$, such coordinates are called wave coordinates. They are in fact a special instance of the so-called generalized wave coordinates, where instead of asking $H = 0$ we allow H to be equal to a non-zero quantity depending on $\mathbf{g}$ and not on $\partial\mathbf{g}$. In the literature, this extra freedom has been used for instance to help determine some spacelike or timelike asymptotic behaviour in stability proofs such as [70,76–78] or even in numerical simulations [79]. In the context of Burnett's conjecture, the role of H is somewhat different. First, the main result of [57] by the author of this review:

Theorem 6 ([57]). *If $(\mathbf{g}_0, F_{\mathbf{A}}, u_{\mathbf{A}})$, with $\mathbf{A}$ belonging to a finite set, is a generic solution on $[0,1]\times\mathbb{R}^3$ of (12) in wave coordinates and close to Minkowski, then there exists a sequence $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ of solutions of the Einstein vacuum equations in generalized wave coordinates on $[0,1]\times\mathbb{R}^3$ and such that $\mathbf{g}_\lambda \longrightarrow \mathbf{g}_0$ uniformly and $\partial\mathbf{g}_\lambda \longrightarrow \partial\mathbf{g}_0$ weakly when λ tends to 0.*

The existence of the background solution $(\mathbf{g}_0, F_{\mathbf{A}}, u_{\mathbf{A}})$ is proved in [80], and the case of a single null dust is considered in [56,81]. Theorem 6 is the strict equivalent of Theorem 5 without any symmetry assumption and the multiphase geometric optics strategy is also followed as $\mathbf{g}_\lambda$ is of the form

$$\mathbf{g}_\lambda = \mathbf{g}_0 + \lambda\sum_{\mathbf{A}}\mathbf{g}^{(1)}_{\mathbf{A}}\left(\frac{u_{\mathbf{A}}}{\lambda}\right) + \lambda^2\left(\sum_{\mathbf{A}}\mathbf{g}^{(2)}_{\mathbf{A}}\left(\frac{u_{\mathbf{A}}}{\lambda}\right) + \sum_{\mathbf{A}\neq\mathbf{B},\pm}\mathbf{g}^{(2,\pm)}_{\mathbf{AB}}\left(\frac{u_{\mathbf{A}}\pm u_{\mathbf{B}}}{\lambda}\right)\right) + \text{remainder}. \tag{18}$$

Note that if we forget about the remainder and consider only one dust, the ansatz (18) is precisely the one considered in the approximate construction [4]. By plugging (18) into (17) and asking for $R_{\alpha\beta}(\mathbf{g}_\lambda) = 0$, we derive a system of transport equations and polarization conditions for the profiles $\mathbf{g}^{(i)}$, the former coming from the truely hyperbolic terms in (18) and the latter from

the gauge terms ∂H. Contrary to the elliptic gauge in $\mathbb{U}(1)$ symmetry and because of the absence of decoupling, the transport equations and polarization conditions must be shown to be compatible at any order in λ. The polarization conditions are of the schematic form

$$\mathbf{P}\left[\mathbf{g}_{\mathbf{A}}^{(1)}|u_{\mathbf{A}}\right]=0,\quad \mathbf{P}\left[\mathbf{g}_{\mathbf{A}}^{(2)}|u_{\mathbf{A}}\right]=\left|\mathbf{g}_{\mathbf{A}}^{(1)}\right|_{\mathbf{g}_0}^2 \mathrm{d}u_{\mathbf{A}}\otimes \mathrm{d}u_{\mathbf{A}},\quad \mathbf{g}_{\mathbf{AB}}^{(2,\pm)}-\mathbf{P}\left[\mathbf{g}_{\mathbf{AB}}^{(2,\pm)}|u_{\mathbf{A}}\pm u_{\mathbf{B}}\right]=\mathbf{g}_{\mathbf{A}}^{(1)}\star\mathbf{g}_{\mathbf{B}}^{(1)}, \tag{19}$$

where the $\star$ denotes some tensor contractions and $\mathbf{P}$ has been defined in (3). The first condition in (19) is the equivalent of Choquet-Bruhat's (4) and the non-trivial RHS's of the last two conditions are due to the lack of null condition of the semilinear terms in (17). The solvability of the last two conditions in (19) is a structural miracle, in some sense equivalent to the ellipticity of the equations for $\mathfrak{g}$ in (15). However, as opposed to [51], here the metrics $\mathbf{g}_\lambda$ don't satisfy the same gauge conditions as their limit $\mathbf{g}_0$, precisely because of the last two conditions in (19), showing that H plays the role of a bin in which we hide undesired terms. Note that the first condition in (19) still implies that

$$H(\mathbf{g}_\lambda)\longrightarrow H(\mathbf{g}_0),\quad \text{uniformly,}\quad \text{and}\quad \partial H(\mathbf{g}_\lambda)\longrightarrow \partial H(\mathbf{g}_0)\quad \text{weakly,} \tag{20}$$

when λ tends to 0.

As explained in Section 3.5, the existence of an exact solution of the form (18) up to time 1 does not follow from any general result. This can be directly seen on the wave equation satisfied by the remainder in (18), which, due to the quasilinearity of the wave operator in (17) is coupled to the transport equation for $\mathbf{g}^{(2)}$, ultimately leading to the apparent loss of one derivative (in consistence with Burnett's regime living one derivative lower than what Theorem 4 provides). Regaining this derivative makes crucial use of the high-frequency ansatz (18).

4.2. *The direct conjecture*

In this section, we review the results in [50,52,53] on the direct conjecture in $\mathbb{U}(1)$ symmetry or in generalized wave coordinates. These results will assume the existence of a sequence of vacuum metrics $(\mathbf{g}_\lambda)_{\lambda\in(0,1]}$ converging in some sense, actually in a stronger compared to (10), to a given metric $\mathbf{g}_0$. Their goal is then to compute the Einstein tensor of $\mathbf{g}_0$ and show that it satisfies a transport equation. The main tool are microlocal defect measures and the structural miracle is compensated compactness.

4.2.1. *Microlocal defect measures*

Microlocal defect measures (sometimes also called H-measures) have been introduced to detect the lack of strong convergence of a sequence in L^2. We refer to [82] for a nice introduction to these measures and to their semi-classical counterparts. Introduced in [83,84], microlocal defect measures are associated to sequences $(u_k)_{k\in\mathbb{N}}$ of functions in $L^2(\mathbb{R}^{d+1},\mathbb{C})$ and converging weakly to 0, and in these references it is shown that there exists a non-negative complex-valued Radon measure μ (the so-called microlocal defect measure) such that

$$\lim_{k\to+\infty}\langle Au_k,u_k\rangle_{L^2(\mathbb{R}^{d+1},\mathbb{C})}=\int_{S^*\mathbb{R}^{d+1}} a(x,\xi)\mathrm{d}\mu, \tag{21}$$

for all order 0 pseudo-differential operator A with principal symbol a homogeneous of degree 0, and where $S^*\mathbb{R}^{d+1}$ is the cosphere bundle of $\mathbb{R}^{d+1}$. For instance, if $u_k(x)=k^{d+1/2}\chi(k(x-x_0))$ (with χ compactly supported) then $\mathrm{d}\mu$ is of the form $\delta_{x_0}\otimes\mathrm{d}\nu$, and if $u_k(x)=\varphi(x)\cos(x\cdot\omega)$ then $\mathrm{d}\mu=|\varphi|^2\mathrm{d}x\otimes\delta_\omega$. These two simple examples illustrate respectively lack of strong convergence by concentration and oscillation and the second one reminds us of the backreaction's dust nature obtained via geometric optics (see Section 4.1). Microlocal defect measures satisfy several important properties, the first of them being localization: if P is an order m differential operator with principal symbol p, and if $(P(u_k))_{k\in\mathbb{N}}$ is relatively compact in H^{-m}_{loc} then $p\mathrm{d}\mu=0$. This shows

that $\mathrm{d}\mu$ is supported (in phase space) on the zero set of the symbol p. In the context of Burnett's conjecture, P will always be a wave operator associated to a Lorentzian metric, and the zero set of p in this case is precisely $\mathscr{P}_0$, the zero mass shell! The second key property of microlocal defect measures in the context of wave equations is that they propagate. This has been first proved in the linear case in [85]. As a toy model, consider a sequence u_k such that ∂u_k and $\Box_{\mathbf{m}} u_k$ (where $\mathbf{m}$ is the Minkowski metric) converge respectively weakly and strongly to 0 in L^2, it can be shown that the microlocal defect measure associated to derivatives of u_k satisfy

$$\left\{\mathbf{m}^{\alpha\beta}\xi_\alpha\xi_\beta, a\right\}\frac{\mathrm{d}\nu}{|\xi|^2} = 0 \tag{22}$$

for all a homogeneous of degree 1, and where $\{\cdot,\cdot\}$ denotes the Poisson bracket in the (x,ξ) variables. The relation (22) is a weak formulation of the Vlasov equation (7).

Note that microlocal defect measures are only sensible to the direction $\xi/|\xi|$ of the lack of strong convergence, and in this regard semi-classical defect measures are more precise. Both types of defect measures are powerful tools with applications in the theory of homogenization but also to the control of wave equations, see for instance [86,87]. However, as mentioned earlier in this text, the phenomenon at the heart of Burnett's conjecture is compensated compactness, the most famous example of which is the div-curl lemma (see [32,33]). Generalized in [83], it stipulates that one can pass to the limit in particular quadratic expressions. For the application to Burnett's conjecture, one would like a similar statement for cubic interactions. Prior to the results discussed in the sections below, the only known example of trilinear compensated compactness was the result [88], where it is shown that $u_k^{(1)}u_k^{(2)}u_k^{(3)}$ converges weakly to $u_\infty^{(1)}u_\infty^{(2)}u_\infty^{(3)}$ if $u_k^{(i)}$ converges weakly to $u_\infty^{(i)}$ in L^2 and if there exists vector fields $X^{(i)}$ satisfying a particular coherence condition and such that $X^{(i)}u_k^{(i)}$ is bounded in L^2.

As geometric optics is the perfect tool to prove Burnett's reverse conjecture, it turns out that microlocal defect measures are the perfect tools to prove Burnett's direct conjecture, as we will now show.

4.2.2. *The* $\mathbb{U}(1)$ *symmetry*

In this section, we describe the work [52] and the subsequent [50], which consider spacetimes enjoying $\mathbb{U}(1)$ symmetry in the elliptic gauge, a framework already described in Section 4.1.2. To state the main theorems of these works, we first define the kinetic spacetimes under consideration. If $\mathrm{d}\nu$ is a non-negative Radon measure on $S^*\mathbb{R}^{2+1}$, we say that $(g,\varphi,\omega,\mathrm{d}\nu)$ is a radially-averaged measure solution to the massless Einstein–Vlasov system in $\mathbb{U}(1)$ symmetry if the first two equations in (14) hold, if for all vector field X smooth and compactly supported we have

$$\int_{\mathbb{R}^{2+1}} R_{\mu\nu}(g)X^\mu X^\mu = \int_{\mathbb{R}^{2+1}}\left(2(X\varphi)^2 + \frac{1}{2}\mathrm{e}^{-4\varphi}(X\omega)^2\right) + \int_{S^*\mathbb{R}^{2+1}} g(\xi,X)^2\frac{\mathrm{d}\nu}{|\xi|^2}, \tag{23}$$

if for all f compactly supported on $\mathscr{M}$ we have

$$\int_{S^*\mathbb{R}^{2+1}} f g^{\alpha\beta}\xi_\alpha\xi_\beta\frac{\mathrm{d}\nu}{|\xi|^2} = 0, \tag{24}$$

and if for all a defined on $T^*\mathbb{R}^{2+1}$ and homogeneous of degree 1 we have

$$\int_{S^*\mathbb{R}^{2+1}}\left(g^{\alpha\beta}\xi_\alpha\partial_\beta a - \frac{1}{2}\partial_\mu g^{\alpha\beta}\xi_\alpha\xi_\beta\partial_{\xi_\mu}a\right)\frac{\mathrm{d}\nu}{|\xi|^2} = 0. \tag{25}$$

Note that, when put together, Equations (23)–(25) are a weak formulation of the massless Einstein–Vlasov system in $\mathbb{U}(1)$ symmetry. The following is the main result of [52] by Huneau and Luk and of [50] by Guerra and Teixeira da Costa:

Theorem 7 ([50,52]). *Let $(g_n,\varphi_n,\omega_n)_{n\in\mathbb{N}}$ a sequence of solutions to (14) in the elliptic gauge such that there exists (g_0,φ_0,ω_0) satisfying the elliptic gauge conditions and such that (g_n,φ_n,ω_n)*

converges to $(g_0, \varphi_0, \omega_0)$ *uniformly and weakly in* $W^{1,p}$ *for some* $p > 2$*, then there exists* $\mathrm{d}\nu$ *a non-negative Radon measure on* $S^*\mathbb{R}^{2+1}$ *such that* $(g_0, \varphi_0, \omega_0, \mathrm{d}\nu)$ *satisfy (23) and (24). If moreover there exists a sequence* $(\lambda_n)_{n\in\mathbb{N}}$ *converging to 0 such that for all compact* $K \subset \mathbb{R}^{2+1}$ *we have*

$$\sum_{k=0}^{q} \lambda_n^{k-1} \left\| \partial^k \left((g_n, \varphi_n, \omega_n) - (g_0, \varphi_0, \omega_0) \right) \right\|_{L^\infty(K)} \leq C_K, \tag{26}$$

with $q = 4$ *in [52] and* $q = 2$ *in [50], then* $(g_0, \varphi_0, \omega_0, \mathrm{d}\nu)$ *satisfy (25).*

Note that the existence of sequences $(g_n, \varphi_n, \omega_n)_{n\in\mathbb{N}}$ as in Theorem 7 and moreover unbounded in $W^{s,p}$ if $s > 1$ (so that the measure $\mathrm{d}\nu$ obtained has a chance to be non-zero) is a consequence of [51]. As it can be seen in Theorem 7, the existence of the measure $\mathrm{d}\nu$ such that $(g_0, \varphi_0, \omega_0, \mathrm{d}\nu)$ satisfy (weakly) the Einstein equations with a massless kinetic stress energy tensor only follows from the Burnett convergence regime (10) (allowing both concentration and oscillation), while the fact that $\mathrm{d}\nu$ solves (weakly) a Vlasov equation requires the stronger high-frequency assumption (26). We now discuss these two parts of Theorem 7 successively, using again our concise notations $(\mathfrak{g}, \psi)$ from Section 4.1.2.

The existence of the measure $\mathrm{d}\nu$ follows from the general existence results [83,84] applied to the sequence $\left(\partial(\psi_n - \psi_0)\right)_{n\in\mathbb{N}}$, which by assumption converge to 0 only weakly. We wish now to pass to the limit in the system (15) that $(\mathfrak{g}_n, \psi_n)$ satisfy. For the wave map system, this is possible thanks to standard bilinear compensated compactness applied to the null form $g^{-1}(\mathrm{d}\psi_n, \mathrm{d}\psi_n)$. For the Ricci equation, the elliptic gauge allows to improve weak convergence to strong convergence for $\nabla\mathfrak{g}_n$, so that we can deal with $(\nabla\mathfrak{g})^2$, and thanks to the equivalent of (21) here, the term $(\partial\psi_n)^2$ converges weakly to $(\partial\psi_0)^2 + \xi^2 \mathrm{d}\nu/|\xi|^2$. The support property of $\mathrm{d}\nu$ follows again from bilinear compactness used together with the localization property of microlocal defect measure and the wave maps system for $\psi_n - \psi_0$. In conclusion, we have proved the first part of Theorem 7.

The most difficult part of [50,52] is to deduce the weak Vlasov equation (25) from the additional high-frequency assumption (26). We repeat that a similar statement was known for linear wave equations, while here a quasilinear hyperbolic-elliptic system is considered and trilinear compensated compactness needs to be uncover. The derivation relies fully on the wave map structure of the equations for ψ, both for the semilinear terms satisfying the null condition and for the coupling between the two components of the wave map, and on the coupling to the elliptic equations for the metric in order to handle quasilinear terms.

4.2.3. *Generalized wave coordinates*

We describe now the work [53], which proves Burnett's direct conjecture in generalized wave gauge. This framework has already been discussed in Section 4.1.3 above. In consistence with Theorem 6 and in particular (20), [53] considers sequences $(\mathbf{g}_n)_{n\in\mathbb{N}}$ of solutions to the Einstein vacuum equations satisfying two sets of assumptions (27) and (28) below, with $\mathbf{g}_0$ a given metric

$$\left\|\mathbf{g}_n - \mathbf{g}_0\right\|_\infty + \left\|H(\mathbf{g}_n) - H(\mathbf{g}_0)\right\|_\infty \leq \lambda_n, \quad \left\|\partial\mathbf{g}_n\right\|_\infty \leq C, \tag{27}$$

$$\lambda \left\|\partial^2\mathbf{g}_n\right\|_\infty + \left\|\partial H(\mathbf{g}_n)\right\|_\infty \leq C, \tag{28}$$

where $C > 0$ and $(\lambda_n)_{n\in\mathbb{N}}$ is a given sequence converging to 0. The following result from [53] is due to Huneau and Luk:

Theorem 8 ([53]). *Let* $(\mathbf{g}_n)_{n\in\mathbb{N}}$ *be a sequence of solutions to the Einstein vacuum equations and* $\mathbf{g}_0$ *a metric. If (27) holds, then there exists a non-negative Radon measure* $\mathrm{d}\mu$ *on* $T^*\mathbb{R}^{3+1}$ *supported on the zero mass shell of* $\mathbf{g}_0$ *and such that*

$$\int_{\mathbb{R}^{3+1}} \psi R_{\mu\nu}(\mathbf{g}_0) = \int_{S^*\mathbb{R}^{3+1}} \psi \xi_\mu \xi_\nu \mathrm{d}\mu$$

for all function ψ smooth and compactly supported. If moreover (28) holds, then for any a homogeneous of degree 1 we have $\{\mathbf{g}_0^{\alpha\beta}\xi_\alpha\xi_\beta, a\}\mathrm{d}\mu = 0$.

Theorem 8 is the equivalent of Theorem 7 in generalized wave coordinates, which in particular removes any symmetry assumption. Its conclusions are the same, and it splits similarly into two parts. First, getting the existence of $\mathrm{d}\mu$ and the Ricci equation for $(\mathbf{g}_0, \mathrm{d}\mu)$ requires only (27) and follows again from [83,84]. More precisely we set $\mathrm{d}\mu := \mathbf{g}_0^{\alpha\rho}\mathbf{g}_0^{\beta\sigma}\left(1/4\mathrm{d}\mu_{\rho\beta\alpha\sigma} - 1/2\mathrm{d}\mu_{\rho\alpha\beta\sigma}\right)$ where each $\mathrm{d}\mu_{\rho\beta\alpha\sigma}$ is associated to $\partial(\mathbf{g}_n - \mathbf{g}_0)$ in the following way

$$\lim_{n\to+\infty}\left\langle \partial_\gamma(\mathbf{g}_n - \mathbf{g}_0)_{\alpha\beta}, A\partial_\delta(\mathbf{g}_n - \mathbf{g}_0)_{\rho\sigma}\right\rangle_{L^2} = \int_{S^*\mathbb{R}^{3+1}} a\xi_\gamma\xi_\delta \mathrm{d}\mu_{\alpha\beta\rho\sigma}.$$

The measure $\mathrm{d}\mu$ is tuned to match the exact expression of the semilinear terms in the Ricci tensor that don't satisfy the null condition. Thanks to the div-curl lemma applied to the null form $\mathbf{g}^{-1}(\mathrm{d}\varphi, \mathrm{d}\psi)$ and to estimates from [89] applied to the null forms $\partial_\mu\varphi\partial_\nu\psi - \partial_\mu\psi\partial_\nu\varphi$ (which are new compared to the $\mathbb{U}(1)$ symmetry setting), one shows that the semilinear terms without the null condition are the only one contributing when passing to the limit in $R_{\mu\nu}(\mathbf{g}_n) = 0$. The proof of the propagation property of $\mathrm{d}\mu$, i.e. the fact that under the stronger assumption (28) we have $\{\mathbf{g}_0^{\alpha\beta}\xi_\alpha\xi_\beta, a\}\mathrm{d}\mu = 0$, requires to compute the weak limit of terms of the form $\left\langle \partial(\mathbf{g}_n - \mathbf{g}_0), A\left(\partial(\mathbf{g}_n - \mathbf{g}_0), \partial(\mathbf{g}_n - \mathbf{g}_0)\right)\right\rangle_{L^2}$. One of the new issue compared to the $\mathbb{U}(1)$ symmetry are the semilinear terms not satisfying the null condition, and trilinear compensated compactness for these terms can be shown using the generalized wave coordinates condition $\|\partial H(\mathbf{g}_n)\|_\infty \leq C$. This shows again that the failure of null condition can be tamed by gauge terms in the Ricci tensor.

5. Angularly regular spacetimes

As explained in Section 3.5, the article [55] deserves its own section because it is very different in spirit from the articles discussed in Section 4 above. In particular, it provides an analytical framework where both parts of Burnett's conjecture can be studied, and in fact the reverse part will be deduced from the direct part.

This framework is the one of angularly regular spacetimes in double null gauge. Requiring no symmetry assumption, the double null gauge has proved very flexible over the years, see for instance the important works [90–92]. If S is a compact 2-surface and $u_*, \underline{u}_* > 0$, the Lorentzian manifold $\left([0, u_*] \times [0, \underline{u}_*] \times S, \mathbf{g}\right)$ is said to be in double null gauge if

$$\mathbf{g} = -4\Omega^2 \mathrm{d}u\mathrm{d}\underline{u} + \gamma_{AB}\left(\mathrm{d}\theta^A - b^A\mathrm{d}u\right)\left(\mathrm{d}\theta^B - b^B\mathrm{d}u\right), \tag{29}$$

where $\Omega > 0$ is a function, γ is a Riemannian metric on S, (θ^1, θ^2) are coordinates on S and b is vector field tangent to S. The local existence of spacetimes in double null gauges is usually formulated with a characteristic initial value problem, where characteristic data are put on two intersecting null hypersurfaces $H_0 = \{0\} \times [0, \underline{u}_*] \times S$ and $\underline{H}_0 = [0, u_*] \times \{0\} \times S$, see [93]. In [94], motivated by the study of impulsive gravitational waves (see also [95]), Luk and Rodnianski prove a low-regularity local existence result in double null gauge where the ∂_u and $\partial_{\underline{u}}$ derivatives of the metric coefficients in (29) are allowed to be only in L^2, while the ∂_{θ^A} derivatives must be more regular than H^1. Compared to Theorem 4, the spacetimes produced by [94] are more regular in the angular directions θ^A but require one derivative less in the null directions u and $\underline{u}$, the higher regularity compensating for the lower regularity. The details of the proof of the main result in [94] are out of the scope of this text, let us just say that the structure of the Einstein vacuum equations is exploited at its fullest and that ellipticity in the angular directions plays again a significant role.

Since in the framework of [94] derivatives in the null directions u and $\underline{u}$ are allowed to be only in L^2, the Burnett convergence regime (10) (and thus Burnett's conjecture itself) is in some sense no longer a low-regularity phenomenon, if we restrict ourself to the situation where derivatives

in the null directions are the only ones lacking strong convergence. This allows for a very soft proof of both sides of the conjecture with the only restriction that the kinetic spacetimes under consideration must be two null dusts propagating in the u and $\underline{u}$ directions. More precisely, there target kinetic spacetimes are angularly regular weak solutions to the Einstein-null dusts system of the form $(\mathbf{g}, \mathrm{d}\nu, \mathrm{d}\underline{\nu})$ where $\mathrm{d}\nu$ and $\mathrm{d}\underline{\nu}$ are two non-negative Radon measures such that

$$\begin{cases} \int_{\mathcal{M}} \left(\mathrm{div}X\mathrm{div}Y - D_\mu X^\nu D_\nu Y^\mu\right) \mathrm{dVol}_{\mathbf{g}} = \int_{\mathcal{M}} (Xu)(Yu)\mathrm{d}\nu + \int_{\mathcal{M}} (X\underline{u})(Y\underline{u})\mathrm{d}\underline{\nu}, \\ \int_{\mathcal{M}} \mathbf{g}^{-1}(\mathrm{d}u, \mathrm{d}\varphi)\mathrm{d}\nu = \int_{\mathcal{M}} \mathbf{g}^{-1}(\mathrm{d}\underline{u}, \mathrm{d}\varphi)\mathrm{d}\underline{\nu} = 0, \end{cases} \tag{30}$$

for all smooth X, Y and φ. Note that the equations in (30) are respectively the weak formulation of the first and second equations in (8) in the case of two null dusts. We gather in the following statement rough versions of the main results of [55] by Luk and Rodnianski:

Theorem 9 ([55]). *The following hold:*

(i) *Any sequence of characteristic angularly regular vacuum initial data satisfying uniform bounds gives rise to a sequence of angularly regular vacuum spacetimes in double null gauge (with a common region of existence). Up to a subsequence, this sequence converges in the sense (10) to an angularly regular weak solution to (30).*

(ii) *Any angularly regular weak solution of (30) in double null gauge can be approximated in the sense (10) by a sequence of angularly regular vacuum spacetimes in double null gauge.*

(iii) *Any characteristic angularly regular initial data for (30) gives rise to a unique angularly regular weak solution of (30).*

The three points of Theorem 9 address respectively Burnett's direct conjecture, Burnett's reverse conjecture and the local theory for angularly regular measure-valued null dusts. As we will see, points (ii) and (iii) are consequences of point (i), so we first discuss point (i). The existence part of point (i) is a direct consequence of the local existence result of [94], as does the existence of the limit, with the help of standard compactness arguments using the uniform bounds provided by [94]. The fact that the limit is an angularly regular weak solution to (30) follows from a careful analysis of each quadratic products $\Gamma^{(1)}\Gamma^{(2)}$ in the structure equations of the double null gauge, where the $\Gamma^{(i)}$'s denote Christoffel symbols in the null frame. It turns out that the only such products for which[3]

$$\text{w-lim}\left(\Gamma^{(1)}\Gamma^{(2)}\right) \neq \left(\text{w-lim}\Gamma^{(1)}\right)\left(\text{w-lim}\Gamma^{(1)}\right)$$

are $|\hat{\chi}|^2$ and $|\underline{\hat{\chi}}|^2$ in the Raychaudhuri equations, where $\hat{\chi}$ and $\underline{\hat{\chi}}$ are the shears of the null hypersurfaces. These products, which fully characterize the lack of null condition for the Einstein vacuum equations in this setting, have the additional property to have a sign and to satisfy elliptic equations (the so-called Codazzi equations), two facts that should not surprise the reader anymore. The Radon measures representing the two dusts are then roughly defined by

$$\mathrm{d}\nu := \text{w*-lim}|\hat{\chi}|^2 - |\text{w-lim}\hat{\chi}|^2, \quad \mathrm{d}\underline{\nu} := \text{w*-lim}|\underline{\hat{\chi}}|^2 - |\text{w-lim}\underline{\hat{\chi}}|^2.$$

Once point (i) of Theorem 9 is proved, points (ii) and (iii) are reduced to initial data statements. Indeed, given a solution of (30), if we are able to explicitly (say with oscillations) approach its data by vacuum data, then point (i) and a uniqueness low-regularity statement implies point (ii). Point (iii) follows from similar arguments, and is the first local theory for measure-valued null dusts, in particular allowing the construction of null dust shell solutions (see [96,97] for their significance).

Appart from the fact that the kinetic spacetimes considered in the framework of angularly regular spacetimes in double null gauge are restricted to only two null dusts, Theorem 9 is much more powerful then the statements in Section 4 and we would like to discuss further some aspects

[3] Here w-lim denotes the weak limit.

of it. In addition to providing a unified treatment of both parts of Burnett's conjecture, Theorem 9 completely drops geometric optics as a tool to construct examples. As a consequence, the target kinetic spacetime can be very rough, while the justification of geometric optics expansions in Section 4.1 relies strongly on the smoothness of the target. Moreover, the target kinetic spacetime can be far from Minkowski, while some smallness is required in Theorems 5 and 6. Lastly, as opposed to Theorems 7 and 8 who needed oscillations in order to get the Vlasov equation, lack of strong convergence by concentration is allowed in Theorem 9.

6. Conclusion

In this section we highlight interesting future directions of research on Burnett's conjecture. The review [11] already offers a very rich list of such directions, and we start by briefly presenting some of them.

Even though they clearly give Burnett's conjecture some validity, the various results presented in Sections 4 and 5 have serious limitations. Most of these results concern kinetic solutions close to Minkowski spacetime, with exception Theorem 9 but which is restricted to two null dusts, and dropping the smallness assumption in $\mathbb{U}(1)$ symmetry or generalized wave coordinates would be very interesting. All of these results are local in time, and one could imagine considering as target kinetic spacetimes a global perturbation of Minkowski as constructed in [98,99] and obtaining a global-in-time Burnett approximation result. Opposite to this scenario, the approximation by vacuum spacetimes of the static spherically symmetric self-gravitating solutions of [100] doesn't seem to be true globally in time, in view of the final state conjecture, but perhaps it is for large times. More generally, given any "exotic" solution to the massless Einstein–Vlasov with particular properties, the question of its approximation by vacuum in the spirit of Burnett's conjecture is an interesting one.

By transforming the early [4] into more analytical theorems, the results of Section 4.1 bring mathematical general relativity closer to the vast geometric optics literature. However they represent only the very first step and many more phenomena remain to be explored, such as caustics, diffractive effects etc. Because of the beautiful transparency properties of the Einstein vacuum equations, one could also try to go beyond Burnett's regime and construct stronger oscillations by concretely lowering the value of p in (13). This would correspond to the strongly nonlinear geometric optics regime, used for instance to prove ill-posedness of some supercritical wave equations by Lebeau [101]. In this spirit, the approximate constructions of large amplitudes gravitational waves in [102,103] are definitely of interest, also since they seem to be stronger short pulses than what exists in the mathematical general relativity literature (see [71,90,104,105]).

We conclude this text by coming back to Yvonne Choquet-Bruhat, in honor of which this text has been conceived. In the 90's, she explored with her collaborators alternative formulations of the Einstein vacuum equations, and in particular higher order formulations built with first or second order derivatives of the Ricci tensor. For instance in [106] they give a fourth-order system (with respect to the metric), equivalent to the Einstein vacuum equations, and show that without any choice of gauge this system still retains a very weak form of hyperbolicity [107]. In Remark 6 of the review [108], Choquet-Bruhat even hints that this fourth-order system would enjoy better structural semilinear properties than the Einstein equations, in link with the relaxed null condition introduced in [42]. She repeats this remark in Chapter VIII Section 4 of [2], sending the reader to Chapter XI where high-frequency waves are studied. However, no mention to this exotic formulation is made in Chapter XI, let alone to its application to high-frequency waves, the nature of which thus remains a mystery.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Acknowledgments

The author would like to thank Thibault Damour, Cécile Huneau and Jonathan Luk for illuminating discussions over the years.

References

[1] Y. Choquet-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), pp. 141–225.

[2] Y. Choquet-Bruhat, *General Relativity and the Einstein Equations*, Oxford Mathematical Monographs, Oxford University Press: Oxford, 2009.

[3] G. A. Burnett, "The high-frequency limit in general relativity", *J. Math. Phys.* **30** (1989), no. 1, pp. 90–96.

[4] Y. Choquet-Bruhat, "Construction de solutions radiatives approchées des équations d'Einstein", *Commun. Math. Phys.* **12** (1969), pp. 16–35. Online at http://projecteuclid.org/euclid.cmp/1103841306.

[5] D. R. Brill and J. B. Hartle, "Method of the self-consistent field in general relativity and its application to the gravitational geon", *Phys. Rev.* **135** (1964), B271–B278.

[6] R. A. Isaacson, "Gravitational radiation in the limit of high frequency. I. The linear approximation and geometrical optics", *Phys. Rev.* **166** (1968), pp. 1263–1271.

[7] R. A. Isaacson, "Gravitational radiation in the limit of high frequency. II. Nonlinear terms and the effective stress tensor", *Phys. Rev.* **166** (1968), pp. 1272–1279.

[8] M. A. H. MacCallum and A. H. Taub, "The averaged Lagrangian and high-frequency gravitational waves", *Commun. Math. Phys.* **30** (1973), no. 2, pp. 153–169.

[9] S. R. Green and R. M. Wald, "New framework for analyzing the effects of small scale inhomogeneities in cosmology", *Phys. Rev. D* **83** (2011), no. 8, article no. 084020.

[10] Y. Choquet-Bruhat, *Une Mathématicienne dans Cet étrange Univers : Mémoires*, Odile Jacob: Paris, 2016.

[11] C. Huneau and J. Luk, "High-frequency solutions to the Einstein equations", *Class. Quantum Gravity* **41** (2024), no. 14, article no. 143002.

[12] A. Einstein, "Näherungsweise integration der feldgleichungen der gravitation", *Sitzungsber. Kgl. Preuss. Akad. Wiss.* **1916** (1916), pp. 688–696.

[13] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, W. H. Freeman and Co.: San Francisco, CA, 1973.

[14] Y. Choquet-Bruhat, "Ondes asymptotiques pour un système d'équations aux dérivées partielles non linéaires", *C. R. Acad. Sci. Paris Sér. A* **264** (1967), pp. 625–628.

[15] Y. Choquet-Bruhat, "Ondes asymptotiques et approchées pour des systèmes d'équations aux dérivées partielles non linéaires", *J. Math. Pures Appl. (9)* **48** (1969), pp. 117–158.

[16] Y. Choquet-Bruhat, "Ondes asymptotiques et approchées pour un système d'équations aux dérivées partielles non linéaires", *C. R. Acad. Sci.* **264** (1967), pp. 625–638.

[17] Y. Choquet-Bruhat, "Ondes asymptotiques et approchées pour un système d'équations aux dérivées partielles non linéaires", *Sémin. Jean Leray* **3** (1969), pp. 1–10. MR:255964. Zbl:0177.36404.

[18] A. M. Anile, *Relativistic Fluids and Magneto-fluids: With Applications in Astrophysics and Plasma Physics*, Cambridge Monographs on Mathematical Physics, Cambridge University Press: Cambridge, 1990.

[19] Y. Choquet-Bruhat, "Approximate radiative solutions of Einstein–Maxwell equations", in *Relativity and Gravitation* (G. Kuper Charles, ed.), Gordon and Breach Science Publishers, Inc: New York, 1971, pp. 81–86. Online at https://www.osti.gov/biblio/4661970.

[20] Y. Choquet-Bruhat and A. Greco, "Ondes gravitationnelles à haute fréquence et interaction avec la matière", *C. R. Acad. Sci., Paris, Sér. II, Fasc. b* **323** (1996), no. 2, pp. 117–124.

[21] Y. Choquet-Bruhat and A. H. Taub, "High-frequency, self-gravitating, charged scalar fields", *Gen. Relativ. Gravit.* **8** (1977), no. 8, pp. 561–571.

[22] Y. Choquet-Bruhat and A. Greco, "High frequency asymptotic solutions of Yang–Mills and associated fields", *J. Math. Phys.* **24** (1983), no. 2, pp. 377–379.

[23] Y. Choquet-Bruhat, "Ondes à haute fréquence pour la gravitation avec termes de Gauss–Bonnet", *C. R. Acad. Sci. Paris Sér. I Math.* **307** (1988), no. 12, pp. 693–696.

[24] Y. Choquet-Bruhat, "High frequency waves for stringy gravity", in *Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity, Part A, B (Teaneck, NJ)* (D. G. Blair, M. J. Buckingham and R. Ruffini, eds.), World Scientific Publishing: Singapore, 1989, pp. 349–361. (Perth, 1988). MR:1056882.

[25] H. Andréasson, "The Einstein–Vlasov system/kinetic theory", *Living Rev. Relativ.* **14** (2011), article no. 4.

[26] Y. Choquet-Bruhat, "Problème de Cauchy pour le système intégro-différentiel d'Einstein-Liouville. (Cauchy problem for the Einstein-Liouville integro-differential system)", *Ann. Inst. Fourier* **21** (1971), no. 3, pp. 181–201. Online at https://eudml.org/doc/74046.

[27] A. D. Rendall, "The Newtonian limit for asymptotically flat solutions of the Vlasov–Einstein system", *Commun. Math. Phys.* **163** (1994), no. 1, pp. 89–112.

[28] B. Le Floch and P. G. Lefloch, "Compensated compactness and corrector stress tensor for the Einstein equations in $\mathbb{T}^2$ symmetry", *Port. Math. (N.S.)* **77** (2020), no. 3–4, pp. 409–421.

[29] L. Tartar, *The General Theory of Homogenization*, Lecture Notes of the Unione Matematica Italiana, Springer-Verlag: Berlin; UMI, Bologna, 2009.

[30] D. Christodoulou, "Global solutions of nonlinear hyperbolic equations for small initial data", *Commun. Pure Appl. Math.* **39** (1986), no. 2, pp. 267–282.

[31] S. Klainerman, "The null condition and global existence to nonlinear wave equations", in *Nonlinear Systems of Partial Differential Equations in Applied Mathematics, Part 1 (Santa Fe, NM, 1984)*, Lectures in Applied Mathematics, American Mathematical Society: Providence, RI, 1986, pp. 293–326.

[32] F. Murat, "Compacite par compensation", *Ann. Sc. Norm. Super. Pisa, Cl. Sci., IV. Ser.* **5** (1978), pp. 489–507. Online at https://eudml.org/doc/83787.

[33] L. Tartar, "Compensated compactness and applications to partial differential equations", in *Nonlinear Analysis and Mechanics: Heriot–Watt Symposium, Vol. 4, Edinburgh 1979*, Research Notes in Mathematics, Pitman Publishing Ltd.: London, 1979, pp. 136–212.

[34] C. De Lellis and L. J. Székelyhidi, "Dissipative continuous Euler flows", *Invent. Math.* **193** (2013), no. 2, pp. 377–407.

[35] C. De Lellis and L. J. Székelyhidi, "Weak stability and closure in turbulence", *Phil. Trans. R. Soc. A* (2022), no. 380, article no. 20210091.

[36] P. Isett, "A proof of Onsager's conjecture", *Ann. Math. (2)* **188** (2018), no. 3, pp. 871–963.

[37] J. Shatah, "Normal forms and quadratic nonlinear Klein-Gordon equations", *Commun. Pure Appl. Math.* **38** (1985), pp. 685–696.

[38] J.-L. Joly, G. Metivier and J. Rauch, "Transparent nonlinear geometric optics and Maxwell-Bloch equations", *J. Differ. Equ.* **166** (2000), no. 1, pp. 175–250.

[39] D. Lannes, "Space time resonances [after Germain, Masmoudi, Shatah]", in *Séminaire Bourbaki Volume 2011/2012 exposés 1043-1058*, Astérisque, no. 352, Société mathématique de France, 2013. Online at http://www.numdam.org/item/AST_2013__352__355_0/.

[40] D. Christodoulou and S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series, Princeton University Press: Princeton, NJ, 1993.

[41] H. Lindblad and I. Rodnianski, "The global stability of Minkowski space–time in harmonic gauge", *Ann. of Math. (2)* **171** (2010), no. 3, pp. 1401–1477.

[42] Y. Choquet-Bruhat, "The null condition and asymptotic expansions for the Einstein equations", *Ann. Phys. (8)* **9** (2000), no. 3–5, pp. 258–266.

[43] H. Lindblad and I. Rodnianski, "The weak null condition for Einstein's equations", *C. R., Math., Acad. Sci. Paris* **336** (2003), no. 11, pp. 901–906.

[44] S. Klainerman, I. Rodnianski and J. Szeftel, "The bounded L^2 curvature conjecture", *Invent. Math.* **202** (2015), no. 1, pp. 91–216.

[45] T. Buchert, "Dark energy from structure: a status report", *Gen. Relativ. Gravit.* **40** (2008), no. 2, pp. 467–527.

[46] E. W. Kolb, S. Matarrese and A. Riotto, "On cosmic acceleration without dark energy", *New J. Phys.* **8** (2006), no. 12, article no. 322.

[47] S. R. Green and R. M. Wald, "Examples of backreaction of small-scale inhomogeneities in cosmology", *Phys. Rev. D* **87** (2013), article no. 124037.

[48] T. Buchert et al., "Is there proof that backreaction of inhomogeneities is irrelevant in cosmology?", *Class. Quantum Gravity* **32** (2015), no. 21, article no. 215021.

[49] S. R. Green and R. M. Wald, *Comments on backreaction*, preprint, 2015, 1506.06452.

[50] A. Guerra and R. T. da Costa, *Oscillations in wave map systems and homogenization of the Einstein equations in symmetry*, preprint, 2021, 2107.00942.

[51] C. Huneau and J. Luk, "High-frequency backreaction for the Einstein equations under polarized $\mathbb{U}(1)$-symmetry", *Duke Math. J.* **167** (2018), no. 18, pp. 3315–3402.

[52] C. Huneau and J. Luk, "Trilinear compensated compactness and Burnett's conjecture in general relativity", *Ann. Sci. Éc. Norm. Supér. (4)* **57** (2024), no. 2, pp. 385–472.

[53] C. Huneau and J. Luk, *Burnett's conjecture in generalized wave coordinates*, preprint, 2024, 2403.03470.

[54] C. Huneau and J. Luk, *High-frequency backreaction for the Einstein equations under* $\mathbb{U}(1)$ *symmetry: from Einstein-dust to Einstein–Vlasov*, 2024. In preparation.

[55] J. Luk and I. Rodnianski, *High-frequency limits and null dust shell solutions in general relativity*, preprint, 2020, 2009.08968.

[56] A. Touati, "Geometric optics approximation for the einstein vacuum equations", *Commun. Math. Phys.* **402** (2023), no. 3, pp. 3109–3200.

[57] A. Touati, *The reverse Burnett conjecture for null dusts*, preprint, 2024, 2402.17530.

[58] D. Christodoulou, "Bounded variation solutions of the spherically symmetric Einstein-scalar field equations", *Commun. Pure Appl. Math.* **46** (1993), no. 8, pp. 1131–1220.

[59] G. Métivier, "The mathematics of nonlinear optics", in *Handbook of Differential Equations* (C. M. Dafermos and M. Pokorný, eds.), Handbook of Differential Equations: Evolutionary Equations, North-Holland: Amsterdam, 2009, pp. 169–313.

[60] J. Rauch, in *Hyperbolic Partial Differential Equations and Geometric Optics*, Graduate Studies in Mathematics, American Mathematical Society: Providence, RI, 2012.

[61] P. D. Lax, "Asymptotic solutions of oscillatory initial value problems", *Duke Math. J.* **24** (1957), pp. 627–646.

[62] L. Garding, T. Kotake and J. Leray, "Uniformisation et développement asymptotique de la solution du problème de Cauchy linéaire, à données holomorphes; analogie avec la théorie des ondes asymptotiques et approchées. (Problème de Cauchy I bis et VI)", *Bull. Soc. Math. Fr.* **92** (1964), pp. 263–361.

[63] P.-Y. Jeanne, "Geometric optics for gauge invariant semilinear systems", *Mém. Soc. Math. Fr., Nouv. Sér.* **90** (2002), pp. vi + 160. Online at https://smf.emath.fr/publications/optique-geometrique-pour-des-systemes-semi-lineaires-avec-invariance-de-jauge.

[64] T. Salvi, *Multi-phase high frequency solutions to Klein–Gordon–Maxwell equations in Lorenz gauge in (3 + 1) Minkowski spacetime*, preprint, 2024, 2407.03554.

[65] J. K. Hunter and J. B. Keller, "Weakly nonlinear high frequency waves", *Commun. Pure Appl. Math.* **36** (1983), pp. 547–569.

[66] J.-L. Joly, G. Metivier and J. Rauch, "Generic rigorous asymptotic expansions for weakly nonlinear multidimensional oscillatory waves", *Duke Math. J.* **70** (1993), no. 2, pp. 373–404.

[67] Y. Choquet-Bruhat and V. Moncrief, "An existence theorem for the reduced Einstein equation", *C. R. Acad. Sci. Paris Sér. I Math.* **319** (1994), no. 2, pp. 153–159. MR: 1288395.

[68] Y. Choquet-Bruhat and V. Moncrief, "Future global in time Einsteinian spacetimes with $U(1)$ isometry group", *Ann. Henri Poincaré* **2** (2001), no. 6, pp. 1007–1064.

[69] C. Huneau, "Constraint equations for 3 + 1 vacuum Einstein equations with a translational space-like Killing field in the asymptotically flat case. II", *Asymp. Analy.* **96** (2016), no. 1, pp. 51–89.

[70] C. Huneau, "Stability of Minkowski space–time with a translation space-like Killing field", *Ann. PDE* **4** (2018), no. 1, article no. 12.

[71] S. Alexakis and N. T. Carruth, *Squeezing a fixed amount of gravitational energy to arbitrarily small scales, in* $\mathbb{U}(1)$ *symmetry*, preprint, 2022, 2205.05526.

[72] J. Luk and M. Van de Moortel, *Nonlinear interaction of three impulsive gravitational waves I: main result and the geometric estimates*, preprint, 2021, 2101.08353. [gr-qc].

[73] J. Luk and M. Van de Moortel, "Nonlinear interaction of three impulsive gravitational waves. II: The wave estimates", *Ann. PDE* **9** (2023), no. 1, article no. 10.

[74] C. Huneau and J. Luk, "Einstein equations under polarized $\mathbb{U}(1)$ symmetry in an elliptic gauge", *Commun. Math. Phys.* **361** (2018), no. 3, pp. 873–949.

[75] A. Touati, "Einstein vacuum equations with $\mathbb{U}(1)$ symmetry in an elliptic gauge: Local well-posedness and blow-up criterium", *J. Hyperbolic Differ. Equ.* **19** (2022), no. 04, pp. 635–715.

[76] A. J. Fang, *Nonlinear stability of the slowly-rotating Kerr-de Sitter family*, preprint, 2021, 2112.07183.

[77] A. J. Fang, *Linear stability of the slowly-rotating Kerr-de Sitter family*, preprint, 2022, 2207.07902.

[78] P. Hintz and A. Vasy, "The global non-linear stability of the Kerr-de Sitter family of black holes", *Acta Math.* **220** (2018), no. 1, pp. 1–206.

[79] F. Pretorius, "Evolution of binary black-hole spacetimes", *Phys. Rev. Lett.* **95** (2005), article no. 121101.

[80] Y. Choquet-Bruhat and H. Friedrich, "Motion of isolated bodies", *Class. Quantum Gravity* **23** (2006), no. 20, pp. 5941–5949.

[81] A. Touati, "High-frequency solutions to the constraint equations", *Commun. Math. Phys.* **402** (2023), no. 1, pp. 97–140.

[82] N. Burq, "Semi-classical measures and defect measures", in *Séminaire Bourbaki. Volume 1996/97. Exposés 820–834*, Société Mathématique de France: Paris, 1997, pp. 167–195. ex (French).

[83] P. Gérard, "Microlocal defect measures", *Commun. Partial Differ. Equ.* **16** (1991), no. 11, pp. 1761–1794.

[84] L. Tartar, " H-measures, a new approach for studying homogenisation, oscillations and concentration effects in partial differential equations", *Proc. R. Soc. Edinb., Sect. A, Math.* **115** (1990), no. 3–4, pp. 193–230.

[85] G. A. Francfort and F. Murat, "Oscillations and energy densities in the wave equation", *Commun. Partial Differ. Equ.* **17** (1992), no. 11–12, pp. 1785–1865.

[86] C. Bardos, G. Lebeau and J. Rauch, "Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary", *SIAM J. Control Optim.* **30** (1992), no. 5, pp. 1024–1065.

[87] N. Burq and P. Gérard, "A necessary and sufficient condition for the exact controllability of the wave equation", *C. R. Acad. Sci., Paris, Sér. I, Math.* **325** (1997), no. 7, pp. 749–752.

[88] J. L. Joly, G. Métivier and J. Rauch, "Trilinear compensated compactness and nonlinear geometric optics", *Ann. Math. (2)* **142** (1995), no. 1, pp. 121–169.

[89] A. D. Ionescu and B. Pausader, *The Einstein–Klein–Gordon coupled system: global stability of the Minkowski solution*, Annals of Mathematics Studies, Princeton University Press: Princeton, NJ, 2022.

[90] D. Christodoulou, *The Formation of Black Holes in General Relativity*, EMS Monographs in Mathematics, European Mathematical Society (EMS): Zürich, 2009.

[91] M. Dafermos, G. Holzegel and I. Rodnianski, "The linear stability of the Schwarzschild solution to gravitational perturbations", *Acta Math.* **222** (2019), no. 1, pp. 1–214.

[92] S. Klainerman and F. Nicolò, *The Evolution Problem in General Relativity*, Progress in Mathematical Physics, Birkhäuser Boston, Inc.: Boston, MA, 2003.

[93] J. Luk, "On the local existence for the characteristic initial value problem in general relativity", *Int. Math. Res. Not.* **2012** (2012), no. 20, pp. 4625–4678.

[94] J. Luk and I. Rodnianski, "Nonlinear interaction of impulsive gravitational waves for the vacuum Einstein equations", *Camb. J. Math.* **5** (2017), no. 4, pp. 435–570.

[95] J. Luk and I. Rodnianski, "Local propagation of impulsive gravitational waves", *Commun. Pure Appl. Math.* **68** (2015), no. 4, pp. 511–624.

[96] S. W. Hawking, "Gravitational radiation from collapsing cosmic string loops", *Phys. Lett. B* **246** (1990), no. 1, pp. 36–38.

[97] R. Penrose, "Naked singularities", in *6th Texas symposium on Relativistic astrophysics. New York, NY, USA, December 18–22, 1972*, New York Academy of Sciences: New York, 1973, pp. 125–134.

[98] L. Bigorgne, D. Fajman, J. Joudioux, J. Smulevici and M. Thaller, "Asymptotic stability of Minkowski space–time with non-compactly supported massless Vlasov matter", *Arch. Ration. Mech. Anal.* **242** (2021), no. 1, pp. 1–147.

[99] M. Taylor, "The global nonlinear stability of Minkowski space for the massless Einstein–Vlasov system", *Ann. PDE* **3** (2017), no. 1, article no. 9.

[100] H. Andréasson, D. Fajman and M. Thaller, "Models for self-gravitating photon shells and geons", *Ann. Henri Poincaré* **18** (2017), no. 2, pp. 681–705.

[101] G. Lebeau, "Nonlinear optics and supercritical wave equation", *Bull. Soc. R. Sci. Liège* **70** (2001), no. 4–6, pp. 267–306.

[102] G. Alì and J. K. Hunter, "Large amplitude gravitational waves", *J. Math. Phys.* **40** (1999), no. 6, pp. 3035–3052.

[103] G. Alí and J. K. Hunter, "Diffractive nonlinear geometrical optics for variational wave equations and the Einstein equations", *Commun. Pure Appl. Math.* **60** (2007), no. 10, pp. 1522–1557.

[104] S. Klainerman and I. Rodnianski, "On the formation of trapped surfaces", *Acta Math.* **208** (2012), no. 2, pp. 211–333.

[105] X. An and J. Luk, "Trapped surfaces in vacuum arising dynamically from mild incoming radiation", *Adv. Theor. Math. Phys.* **21** (2017), no. 1, pp. 1–120.

[106] A. Abrahams, A. Anderson, Y. Choquet-Bruhat and J. W. j. York, "Un système hyperbolique non strict pour les équations d'Einstein", *C. R. Acad. Sci., Paris, Sér. II, Fasc. b* **323** (1996), no. 12, pp. 835–841.

[107] J. Leray and Y. Ohya, "Équations et systèmes non-linéaires, hyperboliques non-stricts", *Math. Ann.* **170** (1967), pp. 167–205. Online at https://eudml.org/doc/161535.

[108] Y. Choquet-Bruhat, *Nonstrict and strict hyperbolic systems for the Einstein equations*, preprint, 2001, gr-qc/0111017.

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mécanique
2025, Vol. 353, p. 1-28
https://doi.org/10.5802/crmeca.273

Review article / *Article de synthèse*

Some aspects of spectral and microlocal methods in Mathematical General Relativity

Quelques aspects des méthodes spectrales et microlocales en relativité générale mathématique

Dietrich Häfner[a]

[a] Université Grenoble Alpes, Institut Fourier, 100 rue des maths, 38610 Gières, France
E-mail: Dietrich.Hafner@univ-grenoble-alpes.fr

Abstract. We review some of the important ideas of spectral and microlocal analysis which have been applied to problems in Mathematical General Relativity over the last decades.

Résumé. Nous passons en revue quelques-unes des idées importantes de l'analyse spectrale et microlocale qui ont été appliquées à des problèmes de relativité générale mathématique au cours des dernières décennies.

Keywords. Spectral theory, Microlocal analysis, General Relativity.

Mots-clés. Théorie spectrale, Analyse microlocale, Relativité générale.

Funding. ANR (ANR-20-CE40-0018-01).

Manuscript received 18 October 2024, accepted 4 November 2024.

1. Introduction

The aim of this review is to give an overview of important spectral and microlocal methods that have been applied to questions in Mathematical General Relativity over the last decades. We will try to explain some of them using the wave equation on the De Sitter Schwarzschild metric as a leading example.

It is hard to say exactly when spectral and microlocal methods appear first in the context of General Relativity. Their application was certainly pushed by the desire of understanding quantum fields on a curved background starting in the 1970s. Right from the beginning the interaction was very rich, not only existing methods were applied in a new context, but also this new context contributed to developing new methods. For example in their celebrated paper *Fourier Integral Operators II* [1] Duistermaat and Hörmander cite the construction of Feynman propagators for the Klein Gordon equation on a general Lorentzian manifold as a motivation. It later turned out that an important condition on states on curved spacetime, the so called Hadamard condition, is naturally formulated in terms of microlocal analysis, see the work of Radzikowski [2]. Today these states have been constructed by pseudodifferential calculus, see Gérard and Wrochna [3].

In the construction of preferred quantum states on a curved background scattering theory plays an important role and a whole program concerning scattering theory on black hole type

backgrounds was initiated by Dimock and Kay in the 1980s, see e.g. [4]. This program was later pushed forward by Bachelot in the 1990s, see e.g. [5] and culminated in his rigorous proof of the Hawking effect in the spherically symmetric setting in 1999 [6]. This picture was later completed by Dappiaggi, Moretti and Pinamonti [7] who showed that the Unruh state which is linked to the Hawking effect fulfills the Hadamard condition. Today our understanding of quantum fields is also very complete for fermions on the Kerr spacetime, see Häfner [8] for the Hawking effect and Gérard, Häfner and Wrochna [9] for the Hadamard property of the corresponding Unruh state. These results rely on classical asymptotic completeness results obtained by Häfner and Nicolas in [10]. For bosons the picture is less complete, but see Klein [11] for the Hadamard property of the Unruh state in the De Sitter Kerr metric. Whereas classical asymptotic completeness results are also known for Klein Gordon fields on the (De Sitter) Kerr spacetime (see Häfner [12], Dafermos, Rodnianski, Shlapentokh-Rothman [13] and Gérard, Georgescu, Häfner [14]), the results of Klein rather use decay estimates as obtained e.g. in Hintz and Vasy [15].

Scattering theory for hyperbolic equations is very much linked to dispersive properties of these equations. The importance of these estimates for non linear problems becomes manifest in the fundamental work on the non linear stability of Minkowski spacetime by Christodoulou and Klainerman [16].

It is very fortunate that the so complicated Einstein equations have explicit solutions like the (De Sitter) Kerr family of solutions describing rotating black holes. However to be sure of the physical importance of these solutions, one has to show that these families are stable as solutions of the Einstein vacuum equations. Again dispersive estimates for the linearized problem play a key role in this context.

Whereas the scattering theory developed in the 1980s and 1990s only showed that the local energy of the underlying equations decays to zero, quantitative decay estimates were now needed. Roughly speaking decay of the solution translates to regularity properties of the underlying resolvent and vice-versa. A particular favorable situation in this context is the case when the resolvent in suitable function spaces has a meromorphic extension in a strip beyond the real axis. The poles of this extension are then called resonances. It turns out that this is the case for the De Sitter Kerr metric.

The mathematical study of resonances for black hole spacetimes was initiated by Bachelot and Motet-Bachlot in 1993 [17]. A bit later Sá Barreto and Zworski [18] succeeded to localize the high frequency resonances of a De Sitter Schwarzschild black hole. Later on Bony and Häfner [19] obtained an expansion of the local propagator of the wave equation on the De Sitter Schwarzschild metric. This expansion in particular gives very precise dispersive estimates. The analysis of Bony and Häfner was extended to include also the horizon by Melrose, Sá Barreto and Vasy [20]. Dyatlov then obtained similar results for the De Sitter Kerr metric [21, 22]. The most general version of this theorem including De Sitter Kerr spacetime for small angular momentum of the black hole is due to Vasy [23]. The work of Vasy is an important breakthrough in this context as it doesn't use the symmetries of Kerr spacetime a part from the time invariance. More generally it develops a robust Fredholm theory for non elliptic problems, that turned out to be crucial in the following. This came out as a kind of surprise since Fredholm theory so far seemed to be reserved for elliptic problems. Not surprisingly, these new methods rapidly also gave interesting results in quantum field theory, see e.g. Vasy and Wrochna [24].

An important obstacle for showing dispersive estimates on black hole spacetimes is trapping that appears in these spacetimes. More precisely to show the resonance expansion theorems one has to control the high energy regime of the resolvent which is directly linked to the trapping. Fortunately the trapping on (De Sitter) Kerr spacetimes is very mild, more precisely it is r-normally hyperbolic. Suitable resolvent estimates in the presence of r-normally hyperbolic trapping have been shown by Wunsch-Zworski in [25] and Dyatlov [26], see also Hintz [27].

The program on dispersive estimates on the De Sitter Kerr family culminated in the proof of the non linear stability of this family for small angular momentum of the black holes by Hintz and Vasy in 2016 [28]. Whereas the restriction to small angular momentum has been removed in the linear framework, see Petersen and Vasy [29], it remains for the non linear framework, because growing modes for the linearized problem can't be excluded so far, but see Hintz [30] for important progress in this direction.

In the context of the Kerr family new difficulties appear. On the linear level, the underlying resolvent can't be extended any more meromorphically beyond the real axis. This is a low frequency problem and very close to what happens when considering evolution equations on an asymptotically euclidean manifold. It turns out that the resolvent has nevertheless certain C^k properties down to the real axis permitting to show polynomial decay for several evolution equations, see e.g. the series of papers by Bony and Häfner [31, 32]. However, for the wave equation the low frequency analysis becomes significantly simpler if one uses a hyperboloidal foliation, see Vasy [33,34]. We also refer to Baskin, Vasy and Wunsch [35,36] for the asymptotically Minkowskian context. Let us also mention the work of Gajic and Warnick [37] and Stucker [38] about the definition of quasinormal modes in this context. The above mentioned low frequency problems are very different from the above mentioned high frequency problems and one of the advantages of spectral and microlocal analysis is that these problems can be very clearly separated and then treated by very different methods.

All this has then been used to show linear stability of the Kerr family for small angular momentum, see Häfner, Hintz and Vasy [39]. Recall that most black holes are rapidly spinning, see e.g. Thorne [40], such that it would be important to remove the restriction to small angular momentum. The very robust Fredholm framework makes that the restriction to small angular momentum only comes from the mode analysis. Indeed all essential features for the analysis like the structure of the horizon or the nature of trapping are the same in the case of large angular momentum (see e.g. Dyatlov [41]). In contrast to the De Sitter Kerr case, mode stability for the Teukolsky equation on the Kerr background is known in the full subextreme range in the Kerr case, see Whiting [42]. This was then used by Andersson, Häfner and Whiting [43] to give a full analysis of the mode solutions for the linearized Einstein equations. As an illustration how the whole analysis can be carried out in the full subextreme case we cite Millet [44] who obtains optimal decay results for the Teukolsky equation on the Kerr background, see also the work of Hintz [45] in this context.

The relatively weak decay for linearized gravity[1] makes the analysis of the non linear problem for the Kerr family more difficult than for the the De Sitter Kerr family. In a series of papers Klainerman–Szeftel [46–48] and Giorgi–Klainerman–Szeftel [49] give a proof of non linear stability of the Kerr family in the case of small angular momentum, see also Shen [50]. We also refer to the work of Dafermos, Holzegel, Rodnianski and Taylor [51] for the non linear stability of the Schwarzschild metric.

Let us also briefly mention the case of negative cosmological constant. If one imposes for example Dirichlet boundary conditions for the Klein Gordon equation on the Anti De Sitter Kerr metric, the trapping becomes more severe. As a consequence the decay for the equation becomes weaker, see e.g. Gannot [52] and Holzegel, Smulevici [53].

Microlocal methods also play an important role in the context of strong cosmic censorship in particular in the case of positive cosmological constant. Considering a (De Sitter) Kerr or (De Sitter) Reissner–Nordström black hole, one can see that there exists a stabilizing effect at the black hole horizon, often called redshift effect and an inverse effect at the Cauchy horizon, called blueshift. It is interesting to see that these redshift and blueshift effects are directly linked

[1] More precisely the solution of the linearized problem will decay to some linearized Kerr plus gauge terms.

to radial point estimates which play an important role in establishing the Fredholm setting for linear equations on the background, see Hintz and Vasy [15]. In the case of zero cosmological constant the asymptotic behavior of the metric at infinity leads to only inverse polynomial decay at the black hole horizon and the blueshift effect at the Cauchy horizon seems to be stronger leading in the end to an instability of the Cauchy horizon, see the work of Dafermos [54] as well as Luk and Oh [55,56]. In the case of positive cosmological constant there is a subtle competition between both effects. For example the question if the solution of the wave equation is H^1 near the cosmological horizon is directly linked to the distance to the real axis of the resonance which is closest to the real axis[2], see Cardoso, Costa, Destounis and Hintz [57]. Unfortunately the exact position of these low frequency resonances is very hard to compute. Nevertheless it is possible that strong cosmic censorship fails for certain black hole parameters. Interestingly, it seems that it is always valid on a quantum level, see the work of Hollands, Wald and Zahn [58]. At first glance it might be surprising that only the resonance position that comes somehow from the analysis outside the black hole determines what is going on on the Cauchy horizon. There is indeed a scattering process between the black hole horizons and the Cauchy horizons and scattering theory for the wave equation has been developed in this context by Kehle and Shlapentokh-Rothman [59] in the spherically symmetric case. Remarkably as already observed by Chandresekhar and Hartle [60] the poles of the corresponding scattering matrix can be computed and they lie too far from the real axis to contribute to the instability although one can give data on the black hole horizons which lead to certain instabilities. Note that asymptotic completeness results inside the black hole can also be obtained by time dependent methods, see the work of Häfner, Nicolas and Mokdad [61] and Mokdad and Provci [62]. However this doesn't give a scattering matrix which is needed to study the behavior in the directions transverse to the Cauchy horizon.

We concentrate in this review on spectral analysis, resolvent analysis and Fourier transform. This however doesn't mean that microlocal analysis cannot be applied in non stationary situations. Indeed in general one will be faced with situations where the operators are not stationary but settle down in some sense to a stationary situation. In this situation the here described analysis is the first of two ingredients for the global analysis of the operator. The second ingredient is a control of regularity of solutions of the non stationary equation and requires in addition other techniques which can however be fully microlocal, see e.g. the recent work of Hintz [63]. We also refer to Hintz and Vasy [64] in this context.

In the above review we have concentrated on the mathematical literature on the subject. There is also a very important physics literature that unfortunately we can't review here. Also it goes without saying that only some of the aspects of microlocal analysis in General Relativity can be treated in this review and inevitable important aspects and important contributions will be missing. We hope that the review gives nevertheless an idea of some of the important ideas that have been developed over the last years.

The review is organized as follows. In Section 2 we review some of the important aspects of the (De Sitter) Kerr family of spacetimes which are used in the microlocal approach. A few of the above mentioned results are presented in some detail in Sections 3, 4 gives some minimal notions of mircolocal and semiclassical analysis. In Section 5 we discuss the wave equation on the De Sitter Schwarzschild metric. Through this relatively simple example we try to explain two of the main ideas for showing decay estimates for hyperbolic equations via microlocal methods. Firstly the Fredholm setting and the underlying estimates (in particular radial point estimates) and secondly the high frequency estimates in the presence of trapping.

[2] There is in fact a zero resonance but which is not included in this discussion.

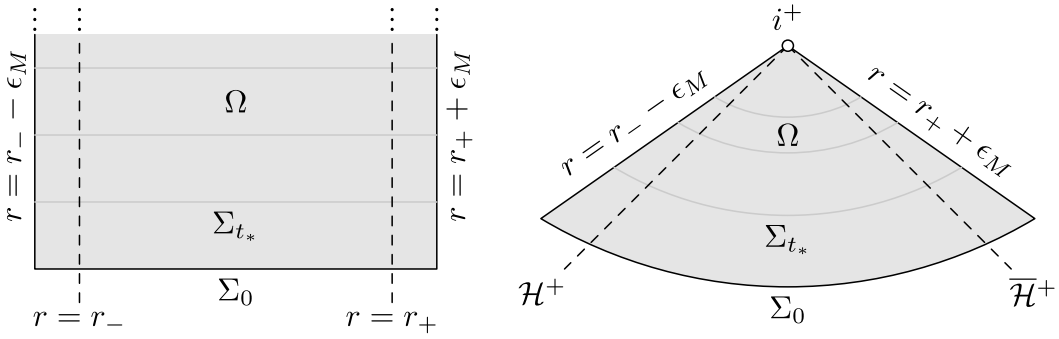


Figure 2.1. Choice of coordinates, source: [28].

2. Black hole spacetimes

2.1. *The De Sitter Schwarzschild metric*

We first consider the De Sitter Schwarzschild metric in Boyer Lindquist coordinates. For $\mathfrak{m} > 0$ we put $b_0 = (\mathfrak{m}, 0)$. The manifold and metric (M, g) are then given by

$$M = \mathbb{R}_t \times \tilde{X}, \quad \tilde{X} = \mathscr{I}_{b_0} \times S^2_\omega, \quad \mathscr{I}_{b_0} = (r_-, r_+), \quad g = F_{b_0}\,\mathrm{d}t^2 - F_{b_0}^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\omega^2, \quad F_{b_0}(r) = 1 - \frac{2\mathfrak{m}}{r} - \frac{\Lambda r^2}{3}.$$

Here $\mathfrak{m} > 0$ is the mass of the black hole and $\Lambda > 0$ is the cosmological constant. We require $9\mathfrak{m}^2\Lambda < 1$ in the following. F_{b_0} then has two positive roots that we call r_-, r_+. The potential $V_{b_0}(r) = F_{b_0}(r)/r^2$ has a unique maximum on (r_-, r_+) given by $r_0 = 3\mathfrak{m}$. The Regge–Wheeler coordinate r_* is defined by the requirement $\mathrm{d}r_*/\mathrm{d}r = F_{b_0}^{-1}$ and that $r_*(3\mathfrak{m}) = 0$. The coordinate $t_*(r_*)$ is defined as a smooth function of r_* such that

$$t_* = \begin{cases} t + r_* & r \le 3\mathfrak{m}, \\ t - r_* & r \ge 4\mathfrak{m}. \end{cases}$$

We find in the coordinates (t_*, r, ω) and for $r \le 3\mathfrak{m}$:

$$g = F_{b_0}\,\mathrm{d}t_*^2 - 2\,\mathrm{d}t_*\,\mathrm{d}r - r^2\,\mathrm{d}\omega^2$$

as well as a similar formula for $r \ge 4\mathfrak{m}$. This shows that the metric extends for some $\epsilon > 0$ to $\mathscr{M} = \mathbb{R}_{t_*} \times X$, $X = (r_- - \epsilon, r_+ + \epsilon) \times S^2_\omega$. The null hypersurfaces $\mathscr{H}^+ = \mathbb{R}_{t_*} \times \{r_-\} \times S^2_\omega$ and $\overline{\mathscr{H}}^+ = \mathbb{R}_{t_*} \times \{r_+\} \times S^2_\omega$ are called the future event horizon and the future cosmological horizon, see Figure 2.1. A similar construction can be performed in the past, constructing then the past event horizon $\mathscr{H}^-$ and the past cosmological horizon $\overline{\mathscr{H}}-$.

2.2. *The De Sitter Kerr metric*

The De Sitter Kerr metric describes black holes with non zero angular momentum. For $\mathfrak{m} > 0$, $\mathfrak{a} \in \mathbb{R}^3$ we put $b = (\mathfrak{m}, \mathfrak{a})$. Let also $a = |\mathfrak{a}|$. We again start with Boyer Lindquist coordinates. For a suitable interval $\mathscr{I}_b \subset \mathbb{R}$ defined below the metric takes the form

$$g_b = \frac{F_b}{(1+\lambda_b)^2\rho_b^2}(\mathrm{d}t - a\sin^2\theta\,\mathrm{d}\phi)^2 - \frac{\kappa_b \sin^2\theta}{(1+\lambda_b)^2\rho_b^2}(a\,\mathrm{d}t - (r^2+a^2)\,\mathrm{d}\phi)^2 - \rho_b^2\left(\frac{\mathrm{d}r^2}{F_b} + \frac{\mathrm{d}\theta^2}{\kappa_b}\right),$$

$$F_b = (r^2+a^2)\left(1 - \frac{\Lambda r^2}{3}\right) - 2\mathfrak{m}r, \quad \rho_b^2 = r^2 + a^2\cos^2\theta, \quad \lambda_b = \frac{\Lambda a^2}{3}, \quad \kappa_b = 1 + \lambda_b\cos^2\theta.$$

Under suitable conditions F_b has two positive roots $r_{b,-}, r_{b,+}$ and $\mathscr{I}_b = (r_{b,-}, r_{b,+})$ in this case. We can construct a similar coordinate t_* as in the De Sitter Schwarzschild case. If one also

changes the coordinate ϕ to some coordinate ϕ^*, then the metric g can be extended in the new coordinates (t_*, r, θ, ϕ^*) to the manifold $\mathscr{M} = \mathbb{R}_{t_*} \times X$, $X = (r_{b,-} - \epsilon, r_{b,+} + \epsilon) \times S^2_{\theta,\phi^*}$. Event horizons and cosmological horizons are defined in a similar way to the De Sitter Schwarzschild case.

2.3. *Schwarzschild and Kerr metric*

In Boyer-Lindquist coordinates the Schwarzschild and the Kerr metric are formally obtained by putting $\Lambda = 0$. Note that the interval $\mathscr{I}_{b_0}$ then becomes $\mathscr{I}_{b_0} = (2\mathfrak{m}, \infty)$ in the Schwarzschild case and the interval $\mathscr{I}_b$ becomes $\mathscr{I}_b = (\mathfrak{m} + \sqrt{\mathfrak{m}^2 - a^2}, \infty)$ in the Kerr case. The spacetimes are now asymptotically Minkowskian, no cosmological horizon exists. We then have $\mathscr{M} = \mathbb{R}_{t_*} \times X$, $X = (\mathfrak{m} + \sqrt{\mathfrak{m}^2 - a^2} - \epsilon, \infty)$. Passing to the conformally rescaled metric $\hat{g} = r^{-2} g$, one can construct future and past null infinity in a similar way to the construction of the cosmological horizons. In the Kerr case we will also need the coordinate $\mathfrak{t}$ which equals t_* near the future event horizon and t near infinity.

2.4. *Some important properties of black hole spacetimes*

2.4.1. *Asymptotic behavior*

The fundamental difference between positive and zero cosmological constant is that the spacetimes are asymptotically De Sitter in the first case and asymptotically Minkowskian in the second case. In the asymptotically Minkowski case we have more precisely:

$$g_{b_0} = g_{(0,0)} + \mathscr{O}(r^{-1}), \quad g_b = g_{b_0} + \mathscr{O}(r^{-2}), \quad r \to \infty.$$

Here $b_0 = (\mathfrak{m}, 0)$, $b = (\mathfrak{m}, \mathfrak{a})$ and $g_{(0,0)}$ is the Minkowski metric. As we will explain in the next section, the different asymptotic behavior has important consequences for the dispersive properties of solutions of the wave equation.

2.4.2. *Superradiance*

It is easy to see that the vector field ∂_t becomes spacelike in a region outside the horizon if the rotation $\mathfrak{a}$ is non zero. This holds for the De Sitter Kerr as well as for the Kerr case. More generally there doesn't exist a global timelike Killing vector field. As a consequence there doesn't exist a positive conserved quantity for the wave equation and the evolution can't be described by a selfadjoint hamiltonian.

2.4.3. *Trapping in the De Sitter Schwarzschild case*

We first consider the De Sitter Schwarzschild case. The associated wave operator is

$$\Box_{g_{b_0}} = F_{b_0}^{-1} \partial_t^2 - \frac{1}{r^2} \partial_r r^2 F_{b_0} \partial_r - r^{-2} \Delta_{\mathbb{S}^2}.$$

Let $P(\sigma) = F_{b_0} \tilde{\Box}_{g_{b_0}}(\sigma)$, where $\tilde{\Box}_{g_{b_0}}(\sigma)$ is the Fourier transform of $\Box_{g_{b_0}}$ with respect to t. Then $P(\sigma)$ has symbol

$$p = \xi^2 + V_{b_0} |\eta|^2 - \sigma^2,$$

where σ, ξ, η are dual to t, r_*, ω and $V_{b_0} = F_{b_0} / r^2$. The projections of the bicharacteristics of p on M are null geodesics, but it is actually more interesting to consider the bicharacteristics themselves. The quantities σ and $|\eta|^2$ are clearly conserved along the evolution and the Hamiltonian equations reduce to

$$\dot{r}_* = 2\xi, \quad \dot{\xi} = -V'_{b_0} |\eta|^2.$$

A solution is clearly given by $(r_*, \xi) = (0,0)$. Linearization around this solution leads to the equation:

$$\dot{r}_* = 2\xi, \quad \dot{\xi} = -V''_{b_0}(0)|\eta|^2 r_*.$$

The matrix

$$\begin{pmatrix} 0 & 2 \\ -V''_{b_0}|\eta|^2 & 0 \end{pmatrix}$$

has eigenvalues $\lambda_\pm = \pm\sqrt{-2V''_{b_0}(0)}|\eta|$. The point (0, 0) is therefore a hyperbolic fixed point of this reduced system.

2.4.4. *r-normally hyperbolic trapping*

In this subsection we describe the trapping arising in black hole spacetimes in a more general framework. We follow closely Dyatlov [26]. Let X be a compact manifold, $p : T^*X \to \mathbb{R}$ a Hamiltonian with associated Hamiltonian vector field H_p. We mimic infinity by an absorption operator Q with semiclassical symbol q, see Section 4.3 for the definition. We define the backward/forward trapped sets $\Gamma_\pm$ and the trapped set K in the following way

(1) $\Gamma_\pm$ are codimension 1 orientable C^∞ submanifolds of T^*X such that $\Gamma_\pm \cap \{p = 0\} \cap \{q = 0\}$ are compact.
(2) if $(x,\xi) \in \{p = 0\} \setminus \Gamma_\pm$, then $\mathrm{e}^{\mp tH_p}(x,\xi) \in \{q > 0\}$ for some $t \geq 0$;
(3) H_p is tangent to $\Gamma_\pm$.
(4) $\Gamma_\pm$ intersect transversely, $K = \Gamma_+ \cap \Gamma_-$ is called the trapped set, and we assume that $WF'_h(Q) \cap K \cap \{|p| \leq \delta\} = \emptyset$ for $\delta > 0$ small enough, see Section 4.3 for a definition of $WF'_h(Q)$.
(5) K is a symplectic codimension 2 submanifold of T^*X.
(6) if $v \in T_K\Gamma_\pm$, then $\mathrm{d}\mathrm{e}^{\mp tH_p} v$ exponentially approaches $TK \subset T_K\Gamma_\pm$ as $t \to \infty$.

We furthermore define $0 < \nu_{\min} \leq \nu_{\max}$ as the maximal and the minimal numbers such that for each $\epsilon > 0$ there exists a constant C such that for each $v \in T_K\Gamma_\pm$

$$C^{-1}\mathrm{e}^{-(\nu_{\max}+\epsilon)t}|\pi(v)| \leq |\pi(\mathrm{d}\mathrm{e}^{\mp tH_p} v)| \leq C\mathrm{e}^{-(\nu_{\min}-\epsilon)t}|\pi(v)|, \quad t \geq 0,$$

where $\pi : T_K\Gamma_\pm \to T_K\Gamma_\pm$ is any fixed smooth linear projection whose kernel is equal to TK. In other words, $\nu_{\min}$ and $\nu_{\max}$ are the minimal and maximal expansion rates in directions transversal to the trapped set. In the above situation we will speak about normally hyperbolic trapping, see Figure 2.2. The notion of r-normally hyperbolic trapping requires in addition that the rate of expansion transversal to K is much larger than the rate of expansion on K. More precisely the trapping is r-normally hyperbolic, if $\nu_{\min} > r\mu_{\max}$, where $\mu_{\max}$ is the maximal expansion rate along K, namely

$$\|\mathrm{d}\mathrm{e}^{tH_p}|_{TK}\| = \mathcal{O}(\mathrm{e}^{(\mu_{\max}+\epsilon)t}), \quad \text{as } |t| \to \infty$$

for all $\epsilon > 0$. It can be shown that the trapping on the (De Sitter) Kerr metric is r-normally hyperbolic for all r, see Dyatlov [41]. This is a structurally stable property, see Fenichel [65] as well as Hirsch, Pugh and Shub [66] for details.

3. Stability results obtained via mircolocal methods

In this section we recall some linear and non linear stability results which have been obtained via microlocal methods.

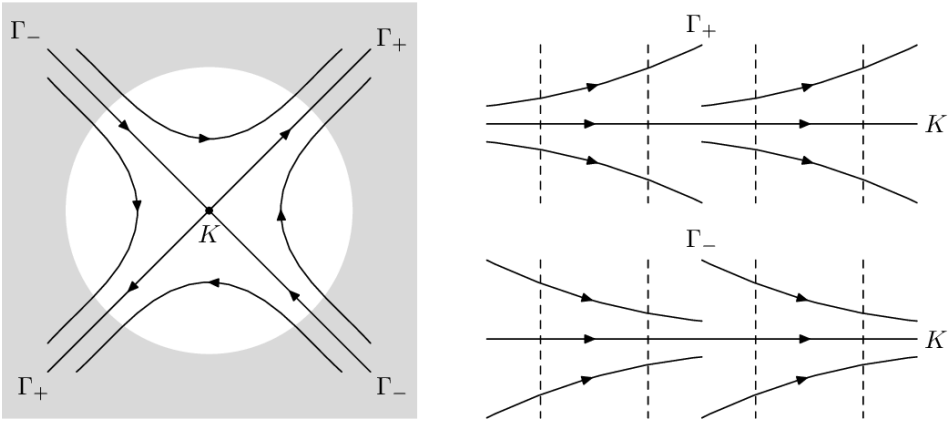


Figure 2.2. Normally hyperbolic trapping, source: [26].

3.1. *The Cauchy problem in General Relativity*

We consider the Einstein vacuum equations

$$\mathrm{Ric}(g) + \Lambda g = 0. \tag{3.1}$$

Initial data for this equation are a triplet (Σ, h, k) consisting of a 3-manifold Σ, a riemannian metric h on Σ and a symmetric 2-tensor k on Σ subject to the *constraint equations*, which are the Gauss–Codazzi equations on Σ implied by (3.1). Let us fix Σ as a submanifold of $\mathscr{M}$. For a Lorentzian metric g on $\mathscr{M}$ we define $\tau(g) = (-g|_\Sigma, \Pi^g_\Sigma)$, where Π^g_Σ is the second fundamental for of Σ within $\mathscr{M}$. A metric g is said to solve the initial value problem with data (Σ, h, k) if

- Σ is spacelike with respect to g;
- $\tau(g) = (h, k)$.

The pioneering work of Choquet-Bruhat [67] shows that the Einstein equations have a local solution for smooth data fulfilling the constraint equations. The non linear stability problem for black hole solutions consists now in slightly perturbing initial data of these solutions and to ask if the initial value problem has a global solution which stays close to one of the members of the family. This is an orbital stability. Note that because of energy radiation each member of the family is not expected to be individually stable.

3.2. *The global nonlinear stability of the De Sitter Kerr family*

The first non linear stability result for black holes was obtained by Hintz and Vasy in 2016 for slowly rotating De Sitter Kerr family. We state it here in its easiest form, see [28] for the full precise formulation of the result. Let us fix a De Sitter Schwarzschild spacetime $(\mathscr{M}, g_{b_0})$. Within this we consider a compact spacelike hypersurface $\Sigma_0 \subset \{t_* = 0\} \subset \mathscr{M}$ extending slightly beyond the event and cosmological horizons. Denote by Σ_{t_*} the translates of Σ_0 by ∂_{t_*} and let $\Omega = \bigcup_{t_* \geq 0} \Sigma_{t_*} \subset \mathscr{M}$ be the spacetime region swept out by these, see Figure 2.1.

Theorem 3.1 (Hintz–Vasy 2016). *Suppose that (h, k) are smooth initial data on Σ_0, satisfying the constraint equations, which are close to the data (h_{b_0}, k_{b_0}) of a De Sitter Schwarzschild spacetime in a high regularity norm. Then there exists a solution g of (3.1) attaining these initial data at Σ_0, and black hole parameters b which are close to b_0, so that*

$$g - g_b = \mathscr{O}(\mathrm{e}^{-\alpha t_*})$$

for a constant $\alpha > 0$ independent of the initial data; that is, g decays exponentially fast to the De Sitter Kerr metric g_b.

Note that data for slowly rotating De Sitter Kerr black holes are included so that the above result is a non linear stability result for the De Sitter Kerr family in case of slow rotation of the black holes.

3.3. *Linear stability of the Kerr family*

As seen above, the decay in the De Sitter Kerr case is very fast. As a consequence passing from the linear to the non linear stability result is not too difficult. This is different in the Kerr case and the full non linear stability of the Kerr family has not been shown so far via spectral and mircolocal methods, but see the series of papers by Klainerman–Szeftel [46], Giorgi–Klainerman–Szeftel [49] for a proof via different methods. However a linear stability result has been obtained by Häfner–Hintz–Vasy via spectral and microlocal methods which we state now in its easiest version. Again we refer to the original version [39] for full details[3]. Consider black hole parameters b close to $b_0 = (\mathfrak{m}_0, 0)$.

Theorem 3.2 (Linear stability of the Kerr family, H-Hintz–Vasy 2019). *Let γ', k' be symmetric 2-tensors on $\Sigma = \mathfrak{t}^{-1}(0)$ satisfying the linearized constraint equations and*

$$|\gamma'| \lesssim r^{-1-\alpha}, \quad |k'| \lesssim r^{-2-\alpha}, \; 0 < \alpha < 1,$$

(and similar bounds for derivatives ∂_ω, $r\partial_r$ up to order 8). Then there exists a symmetric 2-tensor h on $\mathscr{M}$ such that

$$D_{g_b}\mathrm{Ric}(h) = 0, \quad D_{g_b}\tau(h) = (\gamma', k'),$$

which decays to a linearized Kerr metric,

$$h = \dot{g}_b(\dot{b}) + \tilde{h}, \quad |\tilde{h}| \lesssim t_*^{-1-\alpha+}, \quad \left(\dot{g}_b(\dot{b}) := \frac{\mathrm{d}}{\mathrm{d}s} g_{b+s\dot{b}}\Big|_{s=0}\right).$$

Note that Andersson et al. [69] also obtained a linear stability result for the Kerr family, but their decay assumptions on the data are stronger, which eliminates the linearized Kerr solution. Earlier results by Dafermos, Holzegel and Rodnianski [70] as well as Hung, Keller and Wang [71] establish linear stability of the Schwarzschild solution.

3.4. *Remarks on gauge issues*

To carry out the analysis in the linear or non linear case one has to fix a gauge and one usually then considers a gauge fixed operator. On the linear level one can then consider this gauge fixed operator with general initial data. This is an advantage in an iteration scheme to solve the non linear equation because one doesn't have to worry about the linearized constraint equations at each step. On the other hand the gauge fixed linear operator can now have growing modes which are not physical in the sense that they are not solutions of the linearized Einstein equations without gauge. Unfortunately this kind of scenario is produced when one uses the usual wave map gauge. In the De Sitter Kerr case one even expects exponentially growing modes. In the Kerr case the situation is more favorable and only generalized quadratically growing modes can arise. The handling of these gauge issues is beyond the scope of this review. Let us just say that in the Kerr case one can get rid of these modes by the so called constraint damping, we refer to [28] for the De Sitter Kerr case. In the Kerr case once the constraint damping implemented we obtain a linearized operator L_b and the gauge fixed version of Theorem 3.2 is the following.

[3] Note that a small error in [39] has been corrected in [68], the main theorem remains unchanged.

Theorem 3.3 (H-Hintz–Vasy 2019). *Let $\alpha \in (0,1)$, and let $h_0, h_1 \in \mathscr{C}^\infty(\Sigma; S^2 T^*_\Sigma \mathscr{M})$,*

$$|h_0| \lesssim r^{-1-\alpha}, \quad |h_1| \lesssim r^{-2-\alpha}, 0 < \alpha < 1$$

(and similar bounds for derivatives). Let

$$\begin{cases} L_b h = 0, \\ (h|_\Sigma, \mathscr{L}_{\partial_t} h|_\Sigma) = (h_0, h_1). \end{cases}$$

Then there exist $\dot{b} \in \mathbb{R} \times \mathbb{R}^3$ and vector fields $W, V \in$ on $\mathscr{M}$ such that

$$h = \dot{g}'_b(\dot{b}) + \mathscr{L}_V g_b + \mathscr{L}_W g_b + \tilde{h}, \quad |\tilde{h}| \lesssim t_*^{-1-\alpha+}, \quad |\mathscr{L}_W g_b| \lesssim t_*^{-\alpha+}.$$

Here $\dot{g}'_b(\dot{b})$ is a gauge fixed version of $\dot{g}_b(\dot{b})$ and V lies in a fixed 7-dimensional vector space consisting of asymptotic translations, asymptotic boosts and a Coulomb type vector field.

A part from the Coulomb type solutions, the vector field V lies exactly in the space one would expect. In principle one can read off the change of the black hole parameters and the movement of the black hole in the given gauge. Let us also mention that one can get rid of the gauge solution $\mathscr{L}_W g_b$ by admitting a gauge source term.

3.5. *The large a case*

As will be explained later on, the proof of these theorems relies on a very robust Fredholm framework. The only ingredient which is not stable with respect to perturbations is the mode analysis. This mode analysis for the linearized Einstein equations has been carried out by Andersson, Häfner and Whiting [43] also in the large a case. Precise decay estimates for the Teukolsky equation (including the computation of the leading order term) have been obtained by these techniques by Millet in [44]. We give here a rough version of the theorem without defining precisely the Teukolsky operator T_s. We refer to [72] for a review of the geometric background for the Teukolsky operator.

Theorem 3.4 (Millet 2023). *Let $0 \le a < \mathfrak{m}$. We fix $s \in \{0, \pm 1/2, \pm 1, \pm 3/2, \pm 2\}$. Let u_0, u_1* smooth and compactly supported *on Σ_0. The solution u of the Cauchy problem*

$$\begin{cases} T_s u = 0 \\ u(t=0) = u_0 \\ \dfrac{\partial}{\partial t} u(t=0) = u_1 \end{cases}$$

satisfies:

$$|u(r, t_*, \theta, \phi) - \mathfrak{p}_{u_0,u_1}(r, t_*, \theta, \phi)| \le C r^{-1+} t_*^{-2-|s|+s-\epsilon} \left(\frac{t_*}{r} + 1 \right)^{-1-s-|s|}$$

where $\epsilon > 0$. We have

$$\mathfrak{p} = t_*^{-3-2|s|} \frac{(2|s|+2)\left(\frac{t_*}{r}\right)^{2+|s|+s} + 2(|s|-s+1)\left(\frac{t_*}{r}\right)^{1+|s|+s}}{\left(\frac{t_*}{r}+2\right)^{2+|s|+s}} F_{u_0,u_1},$$

where $F_{u_0,u_1}(r,\theta,\phi)$ can be expressed in terms of hypergeometric functions and spin weighted spherical harmonics. The hypotheses of the above theorem can be relaxed. In particular H^N regularity for N sufficiently large is sufficient. Also the compact support of initial data is not necessary, inverse polynomial decay can be permitted and the exact decay rate then depends on the exact decay rate of the initial data. We refer to [44] for details. Similar results also in the large a case have been obtained by Shlapentokh Rothman and Teixeira Da Costa [73, 74] but the precise decay rates are weaker. For entire spin and small a the above result has been obtained before by Ma–Zhang in [75]. The leading order term has already been computed for the wave equation by Hintz [45] in the large a case.

4. Microlocal and semiclassical analysis

In standard microlocal analysis the asymptotic parameter is given by $|\xi|$, where ξ is the fiber variable. We start our representation with a review of that theory. In the semiclassical theory a small parameter h is added to measure the wave length of oscillations. We are then concerned with asymptotics as $h \to 0$, $\xi \to \infty$ and to make the link with standard microlocal analysis h can be thought as $h = 1/|\xi|$. A fundamental reference for the subject is Hörmander [76]. We also refer to Zworski [77] for the semiclassical part. Our presentation owes a lot to the lecture notes of Wunsch [78] and Hintz [79] as well as to the thesis of Hintz [80].

4.1. *Microlocal analysis*

In this section we present some of the main objects of microlocal analysis we will need. Let us start with a differential operator on $\mathbb{R}^n$ of order m,

$$A = \sum_{|\alpha|\le m} a_\alpha(x) D_x^\alpha \in \mathrm{Diff}^m(\mathbb{R}^n).$$

We can rewrite this expression using the Fourier transform $\mathscr{F}$:

$$(Au)(x) = \sum_{|\alpha|\le m} a_\alpha(x)\mathscr{F}^{-1}(\xi^\alpha(\mathscr{F}u)(\xi)) = \frac{1}{(2\pi)^n}\int \mathrm{e}^{\mathrm{i}\langle x,\xi\rangle} \sum_{|\alpha|\le m} a_\alpha(x)\xi^\alpha \hat{u}(\xi)\,\mathrm{d}\xi =: (Op(a)u)(x),$$

where $a(x,\xi) = \sum_{|\alpha|\le m} a_\alpha(x)\xi^\alpha$ is the symbol of the operator. The above writing suggests that one can permit more general symbols $a(x,\xi)$ and then define the corresponding operator $Op(a)$ by

$$(Op(a)u)(x) = \frac{1}{(2\pi)^n}\int \mathrm{e}^{\mathrm{i}x\xi} a(x,\xi)\hat{u}(\xi)\,\mathrm{d}\xi = \frac{1}{(2\pi)^n}\iint \mathrm{e}^{\mathrm{i}\langle (x-y),\xi\rangle} a(x,\xi)u(y)\,\mathrm{d}y\,\mathrm{d}\xi. \tag{4.1}$$

A smooth function a on $T^*\mathbb{R}^n$ is a symbol of order m if the following estimate holds:

$$|\partial_x^\alpha \partial_\xi^\beta a(x,\xi)| \le C_{\alpha\beta}\langle\xi\rangle^{m-|\beta|}, \quad \forall \alpha,\beta \in \mathbb{N}^n.$$

Let $S^m(\mathbb{R}^n)$ be the set of symbols of order m. If $a \in S^m$ we define $Op(a)$ by (4.1). If S^m is the set of symbols of order m, we define the set of pseudodifferential operators of order m as

$$\Psi^m(\mathbb{R}^n) = \{Op(a),\, a \in S^m(\mathbb{R}^n)\}.$$

For symbols in S^m one can define other quantizations, for example the Weyl quantization

$$(a^w(x,D)u)(x) = \frac{1}{(2\pi)^n}\iint \mathrm{e}^{\mathrm{i}\langle (x-y),\xi\rangle} a\Big(\frac{x+y}{2},\xi\Big) u(y)\,\mathrm{d}y\,\mathrm{d}\xi$$

is often used. If $A \in \Psi^m(\mathbb{R}^n)$ we define the principal symbol $\sigma_m(A) \in S^m/S^{m-1}$ by

$$\sigma_m(A) =: a(x,\xi) \bmod S^{m-1}.$$

The principal symbol is independent of the choice of coordinates and it doesn't depend on the particular quantization we use. We now collect some important properties of pseudodifferential operators.

Theorem 4.1 (Fundamental properties of pseudodifferential operators). (1) *Algebra property. Ψ^m is a vector space for each $m \in \mathbb{R}$. If $A \in \Psi^m$ and $B \in \Psi^{m'}$, then $AB \in \Psi^{m+m'}$, $A^* \in \Psi^m$. The composition of operators is associative and distributive. The identity operator is in $\Psi^0(X)$.*

(2) *Boundedness. If $A \in \Psi^m$, then $A \in \mathscr{L}(H^s, H^{s-m})$ for all $s \in \mathbb{R}$.*

(3) *Principal symbol homomorphism.*

$$\sigma_{m+m'}(AB) = \sigma_m(A)\sigma_{m'}(B), \quad \sigma_m(A^*) = \overline{\sigma(A)}.$$

(4) *Symbol of commutator. If $A \in \Psi^m(X)$, $B \in \Psi^{m'}(X)$, then $[A,B] \in \Psi^{m+m'-1}(X)$ and we have*

$$\sigma_{m+m'-1}([A,B]) = \mathrm{i}\{\sigma_m(A), \sigma_{m'}(B)\} = H_a b, \quad \text{if } a = \sigma_m(A), b = \sigma_m(B).$$

Here $\{.,.\}$ is the Poisson bracket and H_a the hamiltonian vector field associated to a.

Definition 4.2. (1) *The symbol $a \in S^m$ is elliptic at a point $(x_0,\xi_0) \in \mathbb{R}^n \times (\mathbb{R}^n \setminus 0)$ if there exists $c, R > 0$ and a conic (in the ξ variable) neighborhood U of (x_0,ξ_0) in $\mathbb{R}^n \times (\mathbb{R}^n \setminus 0)$ such that $|a(x,\xi)| \geq c\langle\xi\rangle^m$ for all $(x,\xi) \in U$, $|\xi| \geq R$. The set of all points at which a is elliptic is called the elliptic set* $\mathrm{Ell}(a)$*, and its complement the characteristic set* $\mathrm{Char}(a)$*. We say that a is uniformly elliptic if it is elliptic at every point with constants R and c uniform in x.*

(2) *The essential support ess supp of $a \in S^m$ is the complement of the set of all $(x_0,\xi_0) \in \mathbb{R}^n \times (\mathbb{R}^n \setminus 0)$ for which $a \in S^{-\infty}$ in a conic neighborhood of (x_0,ξ_0).*

(3) *For $A = Op(a)$ we define its wavefront set as $WF'(A) :=$ ess supp a.*

(4) *Let $u \in \mathscr{D}'(\mathbb{R}^n)$. We define the wave front set of u as follows. (x,ξ) is not contained in $WF(u)$ if and only if there exists an operator $A \in \Psi^0$ which is elliptic at (x,ξ) and a smooth cut-off $\psi \in C_c^\infty(\mathbb{R}^n)$, $\psi(x) \neq 0$, such that $A(\psi u) \in C^\infty(\mathbb{R}^n)$.*

We need two additional important results

Theorem 4.3. *Let $P \in \Psi^m$ be a uniformly elliptic operator. Then there exists $Q \in \Psi^{-m}$ such that*

$$QP - 1 \in \Psi^{-\infty}, \quad PQ - 1 \in \Psi^{-\infty}.$$

Another important result is the sharp Gårding inequality.

Theorem 4.4. *Let $a \in S^{2m+1}$ and $\Re a \geq 0$, then we have for $u \in \mathscr{S}$:*

$$Re\langle Op(a)u, u\rangle \geq -C\|u\|^2_{H^m}.$$

We now radially compactify each fiber of $T^*\mathbb{R}^n$ to obtain $\overline{T^*}\mathbb{R}^n$ and we will denote by $S^*\mathbb{R}^n = \partial\overline{T^*}\mathbb{R}^n$ the cosphere bundle, which we thus view as a boundary at fiber infinity of $\overline{T^*}\mathbb{R}^n$. We will sometimes also need classical symbols. A classical symbol of order m is then $\langle\xi\rangle^m$ times a smooth function on $\overline{T^*}\mathbb{R}^n$. For a classical symbol a we can write down the Taylor series near $1/|\xi| = 0$ to obtain

$$a(x,\xi) \sim \sum_{j=0}^{\infty} a_{m-j}(x,\hat{\xi})|\xi|^{m-j}, \quad a_{m-j} \in C^\infty(\mathbb{R}^n \times \mathbb{S}^{n-1}),$$

and where the tilde denotes an asymptotic expansion-truncating the expansion at the $|\xi|^{m-N}$ term gives an error that is $\mathscr{O}(|\xi|^{m-N-1})$.

Suppose that $A \in \Psi^m$ has compactly supported Schwartz kernel supported in $U \times U$. Then under a coordinate change $\kappa : U \to V$, the principal symbol of an operator $a = \sigma_m(A)$ transforms as a function on the cotangent bundle and the Hamiltonian vector field is invariantly defined under change of coordinates. This allows to define pseudodifferential calculus on a compact manifold X and the above properties remain valid. In the non compact case, some care has to be taken for example by imposing some control at infinity or by requiring that the operators are properly supported.

4.2. *Propagation of singularities*

Let $P \in \Psi^m(X)$ be a classical operator with real homogeneous principal symbol $p_m = \sigma_m(P) \in S^m_{\mathrm{hom}}(T^*X)$. Fix a boundary defining function ρ of fiber infinity in $\overline{T^*}X$. We rescale the Hamiltonian vector field $\mathbf{H}_{p_m} = \rho^{m-1}H_{p_m} \in \mathscr{V}(S^*X)$ and also rescale the principal symbol $\mathbf{p} = \rho^m p_m \in C^\infty(S^*X)$. Within the characteristic set $\mathrm{Char}(P) = \mathbf{p}^{-1}(0)$, the rescaled vector field induces a flow which is merely a rescaling of the Hamiltonian flow of p_m if we identify the subset $\mathrm{Char}(P) \subset S^*X$ with the corresponding conic set of $T^*X \setminus 0$. We remark that H_{p_m} vanishes at a point $\zeta \in S^*X$ if

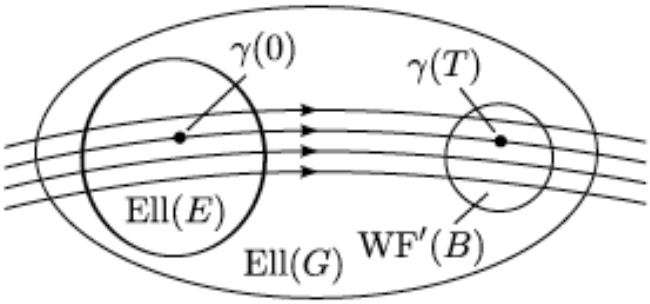


Figure 4.1. Propagation of singularites, source: [80].

and only if H_{p_m} is radial at ζ (i.e. at the ray in $T^* \setminus 0$ associated to ζ), and in this case the integral curve through ζ is constant.

We suppose in the following that $\zeta_0 \in S^*X$ is such that $\mathrm{H}_{p_m}|_{\zeta_0} \neq 0$, hence the hamiltonian flow in $\mathrm{Char}(P)$ starting at ζ_0 is non trivial; denote by $\gamma : [0,T] \to S^*X$ a segment of a null bicharacteristic, i.e. an integral curve of $\mathbf{H}_{p_m}$ in $\mathrm{Char}(P)$ starting at $\gamma(0) = \zeta_0$. The following theorem was first proven by Duistermaat and Hörmander in [1] (Figure 4.1 illustrates this theorem):

Theorem 4.5. *Let $s, N \in \mathbb{R}$. Let $E, B, G \in \Psi^0$. Assume that E is elliptic at ζ_0, B is elliptic at $\gamma(T)$, and G is elliptic at $\gamma([0,T])$, such that every backward null bicharacteristic starting at a point in $WF'(B)$ reaches* $\mathrm{Ell}(E)$ *in finite time. Then*

$$\|Bu\|_{H^s} \leq C(\|GPu\|_{H^{s-m+1}} + \|Eu\|_{H^s} + \|u\|_{H^N}).$$

Remark 4.6.
(1) *The estimates holds in the strong sense that if $u \in \mathscr{E}'(X)$ is such that the right hand side is finite, then so is the left hand side and the estimate holds. The theorem shows that H^s regularity propagates along null bicharacteristics in the following sense. Assuming that we have microlocal H^s control at $\gamma(0)$, we conclude microlocal H^s control at $\gamma(T)$ provided Pu stays in H^{s-m+1} along the way.*
(2) *The hypothesis excludes radial points, we come back to radial points in the special setting of the De Sitter Schwarzschild metric in Section 5.3.2.*

Sketch of the proof. We sketch the proof in the case when P is selfadjoint, $P = P^*$. We assume further $Pu = 0$. We can suppose $T = 1$. We begin by straightening out the flow of $\mathbf{H}_{p_m}$. We introduce local coordinates $(q_1, q') \in \mathbb{R} \times \mathbb{R}^{2n-2}$ on S^*X near ζ_0 such that $\zeta_0 = (0,0)$ and $\mathbf{H}_{p_m} = \partial_{q_1}$. Let for $\delta > 0$, $U_\delta = \{|q_1|, |q'| < \delta\}$. We start with a formal computation and discuss regularity issues later. For a commutant $A = Op(a) \in \Psi^{2s-m+1}$ such that $A = A^*$ to be chosen later we compute:

$$0 = 2\Im\langle Pu, Au\rangle = \langle \mathrm{i}(PA - AP)u, u\rangle = \langle \mathrm{i}[P,A]u, u\rangle. \tag{4.2}$$

Let $\mathbf{a} = \rho^{2s-m+1} a \in C^\infty(S^*X)$. We claim that we can arrange

$$\mathbf{H}_{p_m}\mathbf{a} = -\mathbf{b}^2 - M\mathbf{a} + \mathbf{e}, \tag{4.3}$$

where $\mathbf{b} \in C^\infty(S^*X)$ is non negative, and positive in $\{q_1 \in (\delta, 1+\delta), |q'| < \delta\}$, while $\mathbf{e} \in C^\infty(S^*X)$ is supported in the neighborhood U_δ and $M > 0$ can be chosen large. We define $\mathbf{a}$ as a product $\mathbf{a} = \mathbf{a}_1(q_1)\mathbf{a}'(q')$, where $\mathbf{a}_1$ and $\mathbf{a}'$ are constructed in the following way. Let

$$\psi \in C_0^\infty(\mathbb{R}^{2n-2}),\ \mathrm{supp}\,\psi \subset \{|q'| \leq 2\delta\},\ \psi(q') = 1 \quad \text{for } |q'| < \delta,$$
$$\chi_1 \in C^\infty(\mathbb{R}),\ \mathrm{supp}\,\chi_1 \subset (-\delta, \infty),\ \mathrm{supp}(1-\chi_1) \subset (-\infty, \delta) \text{ (turn off function)},$$

$$\chi_0(q_1) = \begin{cases} \mathrm{e}^{-\frac{F}{1+\delta-q_1}} & q_1 < 1+\delta, \\ 0 & q_1 \geq 1+\delta, \end{cases}$$
$$\mathbf{a}_1(q_1) = \chi_0(q_1)\chi_1^2(q_1), \quad \mathbf{a}'(q') = \psi^2(q').$$

It is not difficult to check that (4.3) can be achieved for $M > 0$ given if we choose F sufficiently large. Going back to a we compute

$$H_{p_m} a = \rho^{-m+1}\mathbf{H}_{p_m}(\rho^{-2s+m-1}\mathbf{a}) = \rho^{-2s}(\mathbf{H}_{p_m}\mathbf{a} - (2s-m+1)\mathbf{a}\rho^{-1}\mathbf{H}_{p_m}\rho). \tag{4.4}$$

Noting that $\rho^{-1}\mathbf{H}_{p_m}\rho$ is bounded and choosing M in (4.3) sufficiently large we see that

$$H_{p_m} a = -b^2 + e - M\tilde{a} + r\tilde{a}, \tag{4.5}$$

where $b = \rho^{-s}\mathbf{b}, e = \rho^{-2s}\mathbf{e}, \tilde{a} = \rho^{-m+1}a, r \in S^0$. Let $\tilde{A}$ be a quantization of $\tilde{a}$ and suppose that we arrange $\tilde{A} = C^*C$. Then $-M\langle \tilde{A}u, u\rangle + \langle Op(r\tilde{a})u, u\rangle \leq C\|u\|^2_{H^{s-(1/2)}}$. Let $B \in \Psi^s(X)$ and $E \in \psi^{2s}(X)$ be quantizations of b and e, respectively. We then obtain:

$$\langle B^*Bu, u\rangle \leq \langle Eu, u\rangle + C\|u\|^2_{H^{s-\frac{1}{2}}}. \tag{4.6}$$

It has to be pointed out that the integration by parts above are not justified at the level of regularity supposed in the hypothesis. The above argument therefore needs a regularization procedure and this can be done for $u \in H^{s-1/2}$ microlocally near $\gamma([0,T])$, see [79] for details. Supposing this done, then the estimate (4.6) implies $Bu \in L^2(X)$, therefore $u \in H^s$ microlocally near $\gamma([0,T])$. Thus starting with the a priori knowledge $u \in H^N(X)$, one can iteratively improve the regularity by $1/2$ in each step, until one obtains H^s-regularity as desired. □

4.3. *Semiclassical analysis*

We briefly introduce some elements of semiclassical analysis that we will need in the following. We refer to Zworski [77] for details. We start by introducing the algebra $\Psi_h^m(X)$ of *semiclassical* pseudodifferential operators, depending on a parameter h. The corresponding symbols $a(x,\xi;h)$ (denoted $a \in S_h^m(X)$) satisfy $a(\cdot,\cdot;h) \in S^m(X)$ uniformly in h as $h \to 0$, with $S^m(X)$ defined in Section 4.1. The semiclassical symbol of $A \in \Psi_h^m$, denoted $\sigma_h(A)$, lies in the space $S_h^m(X)/hS_h^{m-1}(X)$.

We will be concerned in particular with the class $\Psi_h^0(X)$. We moreover require that symbols of the elements of $\Psi_h^0(X)$ are classical in the sense that they have an asymptotic expansion in non negative powers of h. More precisely we require that $a \in \Psi_h^0$ writes

$$a \sim \sum_{j=0}^{\infty} h^j a_j, \text{ meaning } \left|\partial_x^\alpha \partial_\xi^\beta \left(a - \sum_{j=0}^{N-1} h^j a_j\right)\right| \leq C_{\alpha,\beta} h^N \langle\xi\rangle^{-N-|\beta|}$$

for all multiindices α, β and $N \in \mathbb{N}$. For $A, A' \in \Psi_h^0(X)$ with principal symbols a, a' we have

$$\sigma_h(A \circ A') = aa', \quad \sigma_h(A^*) = \bar{a}, \quad \sigma_h(\mathrm{i}h^{-1}[A, A']) = \{a, a'\} = H_a a'.$$

In the context of semiclassical analysis we will mostly use the Weyl quantization:

$$a^w(x, hD)u(x) = \frac{1}{(2\pi h)^n} \iint a\left(\frac{x+y}{2}, \xi\right) \mathrm{e}^{\frac{\mathrm{i}}{h}\langle x-y,\xi\rangle} u(y)\,\mathrm{d}y\,\mathrm{d}\xi.$$

Moreover each $A = a^w(x, hD) \in \Psi_h^0(X)$ acts $L^2(X) \to L^2(X)$ with norm uniformly bounded in h. In fact we have the bound [77, Theorem 13.13]

$$\limsup_{h\to 0} \|A\|_{L^2(X)\to L^2(X)} \leq \sup_{T^*X} |\sigma_h(A)|. \tag{4.7}$$

Each $A \in \Psi_h^k(X)$ has a semiclassical wave front set $WF_h'(A)$, a closed (and not necessarily conic) subset of the fiber-radially compactified cotangent bundle $\overline{T^*}X$; a point $(x,\xi) \in \overline{T^*}X$ does not lie in $WF_h'(A)$ if and only if the full symbol of A satisfies $a(x',\xi') = \mathcal{O}(h^\infty\langle\xi'\rangle^{-\infty})$ for h small

enough and $(x',\xi')\in T^*X$ in a neighborhood of (x,ξ) in $\overline{T^*}X$. We will mostly use the subalgebra of *compactly microlocalized* pseudodifferential operators $\Psi_h^{\mathrm{comp}}(X)\subset\Psi_h^0(X)$. For $A\in\Psi_h^{\mathrm{comp}}(X)$ the wave front set is a compact subset of T^*X. We say that $A=B+\mathcal{O}(h^\infty)$ microlocally in some set $U\subset T^*X$ if $WF_h'(A-B)\cap U=\emptyset$. We will often consider sequences $u(h_j)\in L^2(X)$, where $h_j\to 0$ is a sequence of positive numbers; we typically suppress the dependence on j and write $u=u(h)$. For such a sequence, we say that $u=\mathcal{O}(h^\delta)$ microlocally in some set $U\subset T^*X$, if $\|Au\|_{L^2}=\mathcal{O}(h^\delta)$ for each $A\in\Psi_h^{\mathrm{comp}}(X)$ such that $WF_h'(A)\subset U$. The notion $u=o(h^\delta)$ is defined similarly.

4.4. *Semiclassical defect measures*

Definition 4.7. *Assume that $h_j\to 0$ and $u=u(h_j)\in L^2(X)$ is uniformly bounded as $j\to\infty$. We say that u converges to a non-negative Radon measure μ on T^*X if for each $A=A(h)\in\Psi_h^{\mathrm{comp}}(X)$, we have*

$$\langle A(h_j)u(h_j),u(h_j)\rangle\to\int_{T^*X}\sigma_h(A)\,\mathrm{d}\mu\quad\text{as } j\to\infty. \tag{4.8}$$

See [77, Chapter 5] for an introduction to defect measures, in particular,

(1) *for each sequence $u(h_j)$ uniformly bounded in L^2, there exists a subsequence h_{j_k} such that $u(h_{j_k})$ converges to some μ [77, Theorem 5.2];*
(2) *we have $\mu(U)=0$ for some open set $U\subset T^*X$ if and only if $u=o(1)$ microlocally on U.*

The following basic properties of microlocal measures are easy to prove, see e.g. [26].

Lemma 4.8 (Ellipticity). *Take $P\in\Psi_h^0(X)$ and denote $p=\sigma_h(P)$. Assume that $u=u(h_j)$ converges to some measure μ and $Pu=o(1)$ microlocally in some open set $U\subset T^*X$. Then $\mu(U\cap\{p\neq 0\})=0$.*

Lemma 4.9 (Propagation). *Take $P,W\in\Psi_h^0$, denote $p=\sigma_h(P)$, $w=\sigma_h(W)$, and assume $P^*=P$. Assume that $u=u(h_j)$ converges to some measure μ and denote $f:=(P-\mathrm{i}hW)u$. Then for each $a\in C_0^\infty(T^*X)$ and for each $Y\in\Psi_h^{\mathrm{comp}}(X)$ such that $Y=1+\mathcal{O}(h^\infty)$ microlocally in a neighborhood of* $\operatorname{supp}a$ *we have,*

$$\left|\int_{T^*X}(H_p-2\Re w)a\,\mathrm{d}\mu\right|\le 2\|a\|_\infty\limsup_{h\to 0}(h^{-1}\|Yf\|_{L^2}\|Yu\|_{L^2}). \tag{4.9}$$

In particular, if $f=o(h)$ microlocally in a neighborhood of $\operatorname{supp}a$*, then*

$$\int_{T^*X}(H_p-2\Re wa)\,\mathrm{d}\mu=0.$$

5. The wave equation on the De Sitter Schwarzschild metric

In this section we explain the general strategy for showing decay estimates via microlocal methods for a simple example: the wave equation on the De Sitter Schwarzschild metric. Let $\mathcal{M}=\mathbb{R}_{t_*}\times X$, $X=(r_--\epsilon,r_++\epsilon)\times\mathbb{S}^2_\omega$. Near $r=r_-$ resp. $r=r_+$ the metric is given by

$$g=F\,\mathrm{d}t_*^2-2\,\mathrm{d}t_*\,\mathrm{d}r-r^2\,\mathrm{d}\omega^2\quad\text{resp.}\quad g=F\,\mathrm{d}t_*^2+2\,\mathrm{d}t_*\,\mathrm{d}r-r^2\,\mathrm{d}\omega^2,$$

where $F(r)=1-(2\mathfrak{m}/r)-(1/3)\Lambda r^2$. Recall that the surface gravities are given by $\kappa_\pm=(1/2)F'(r_\pm)$. The wave operator is given near $r=r_-$ resp. $r=r_+$ by

$$\Box_g=2D_{t_*}D_r+\frac{1}{r^2}D_r r^2FD_r-r^{-2}\Delta_{\mathbb{S}^2}\quad\text{resp.}\quad\Box_g=-2D_{t_*}D_r+\frac{1}{r^2}D_r r^2FD_r-r^{-2}\Delta_{\mathbb{S}^2}.$$

Fourier transforming this with respect to t_*, i.e. putting

$$\hat{\Box}_g(\sigma)=\mathrm{e}^{\mathrm{i}t_*\sigma}\Box_g\mathrm{e}^{-\mathrm{i}t_*\sigma}.$$

then gives

$$\hat{\Box}_g(\sigma)=-2\sigma D_r+\frac{1}{r^2}D_r r^2FD_r-r^{-2}\Delta_{\mathbb{S}^2}\quad\text{resp.}\quad\hat{\Box}_g(\sigma)=2\sigma D_r+\frac{1}{r^2}D_r r^2FD_r-r^{-2}\Delta_{\mathbb{S}^2}.$$

We will consider in the following the forward problem

$$\Box_g u = f, \tag{5.1}$$

where $f \in C_0^\infty(\mathscr{M})$ and we require that $\operatorname{supp} u \subset \{t_* \geq -T\}$ for some $T > 0$. Let $\bar{H}^s(X)$ resp. $\dot{H}^s(X)$ be the Sobolev spaces based on extendible resp. supported distributions. Initial value problems can easily be transformed into forward problems and vice versa, see e.g. [44]. The existence of a forward solution is assured by the global hyperbolicity of the spacetime.

5.1. *A key theorem*

A first important result is that the Fourier transformed wave operators form an analytic family of Fredholm operators. Let $\gamma = \max\{(1/\kappa_-), (1/\kappa_+)\}$. We have

Theorem 5.1. *Let $s \geq 1/2$, $\mathscr{X}^s = \{u \in \bar{H}^s(X); \hat{\Box}_g(0)u \in \bar{H}^{s-1}(X)\}$, $\mathscr{Y}^{s-1} = \bar{H}^{s-1}(X)$. The operators*

$$\hat{\Box}_g(\sigma) : \mathscr{X}^s \to \mathscr{Y}^{s-1}$$

form an analytic family of Fredholm operators of index 0 for all $\sigma \in \mathscr{C}$ such that

$$\Im\sigma > \frac{1-2s}{2\gamma}. \tag{5.2}$$

Moreover $\hat{\Box}_g(\sigma)$ is invertible if $\Im\sigma > 0$.

Theorem 5.2. *Let $t_0 \in \mathbb{R}$. There are $C, \nu > 0$ such that for $0 < \delta < C$ and $s > (1/2) + \gamma\delta$ any solution to (5.1) with $f \in \mathrm{e}^{-\delta t_*}\bar{H}^{s-1+\nu}(M)$ and with $\operatorname{supp}(u) \cup \operatorname{supp}(f) \subset \{t_* > t_0\}$ has an asymptotic expansion*

$$u(t_*, r, \omega) - \sum_{j=1}^{N}\sum_{k=0}^{k_j} t_*^k \mathrm{e}^{-\mathrm{i}\sigma_j t_*} a_{j,k}(r,\omega) \in \mathrm{e}^{-\delta t_*}\bar{H}^s(M), \tag{5.3}$$

where $\sigma_1, \ldots, \sigma_N$ are the (finitely many) quasinormal mode frequencies with

$$\Im\sigma_j > -\delta$$

and k_j their multiplicities, and where $\mathrm{e}^{-\mathrm{i}\sigma_j t_*} a_{k,j}$ are the C^∞ generalized quasinormal modes with frequency σ_j.

Remark 5.3.

(1) Quasinormal mode frequencies are also called resonances, they correspond to those σ such that $\hat{\Box}_g(\sigma)$ has a non trivial kernel. By analytic Fredholm theory one obtains a meromorphic family of operators $\hat{\Box}_g^{-1}(\sigma) : \mathscr{Y}^{s-1} \to \mathscr{X}^s$. Resonances or quasinormal mode frequencies are then the poles of this family of operators. These poles may be of higher order which gives rise to the t_*^k coefficients.

(2) Note that close to the black hole horizon, $\hat{\Box}_g(\sigma) = \mathrm{e}^{\mathrm{i}\sigma r_*}\tilde{\Box}_g(\sigma)\mathrm{e}^{-\mathrm{i}\sigma r_*}$, where $\tilde{\Box}_g(\sigma)$ is the Fourier transformed operator with respect to t. Therefore v is mode solution for $\hat{\Box}_g(\sigma)$ ($\hat{\Box}_g(\sigma)v = 0$) if and only if $u = \mathrm{e}^{-\mathrm{i}\sigma r_*}v$ is mode solution for $\tilde{\Box}_g(\sigma)$ and the requirement that v is smooth at the horizon translates to the usual outgoing condition $u \sim \mathrm{e}^{-\mathrm{i}\sigma r_*}$, $r_* \to -\infty$ for u.

(3) The analogue of Theorem 5.1 is false in the Schwarzschild case. Nevertheless the resolvent has good regularity properties down to the real axis, see Vasy [33, 34]. This can then be used to show polynomial decay estimates, see Häfner, Hintz, Vasy [39], Hintz [45] and Millet [44].

(4) There exists a generalized version for the De Sitter Kerr metric (see Vasy [23], Petersen-Vasy [29]), but a priori nothing is said about the existence of resonances with strictly positive imaginary part. The non existence of mode solutions with positive imaginary part of σ has been shown by Whiting in the whole subextreme case for the Kerr solution.

The analogue question for the De Sitter Kerr case remains open, but see Hintz [30] for results in this direction.

(5) High frequency resonances (meaning for large real part of σ) are well approximated by so called pseudopoles which can be computed and which are given by

$$\frac{\sqrt{1-9\Lambda\mathfrak{m}^2}}{3^{\frac{3}{2}}\mathfrak{m}}(\pm(\ell+1/2)-\mathrm{i}(1/2+n)), \quad \ell\in\mathbb{N}, n\in\mathbb{N}_0.$$

These resonances are very closely linked to the trapping of the hamiltonian flow.

(6) As $\Box_g(0)1=0$, 0 is a resonance, in particular the solution *does not* decay. However it can be shown (see [19]) that this is the only real resonance. Together with the high frequency analysis of Section 5.4 this gives the existence of a strip beyond the real axis in which the only resonance is zero. As a corollary one obtains the existence of a constant c with

$$u(t_*,r,\omega)-c\in \mathrm{e}^{-\delta t_*}\bar{H}^s(M).$$

The first version of Theorem 5.2 was shown by Bony and Häfner in [19]. This was formulated in the original Boyer Lindquist coordinates and excluded the horizons. A generalization of this including the horizons was subsequently obtained by Melrose, Sá-Barreto and Vasy in [20]. Later on the theorem was generalized by Dyatlov to the De Sitter Kerr case in [21,22]. The most general version of this was obtained by Vasy in [23] first in the small a case and later generalized to the large a case in collaboration with Petersen [29].

In the remaining of this section we will explain some of the principal ideas in the proof of Theorems 5.1 and 5.2. Since $\mathscr{M}$ is globally hyperbolic there exists a unique forward solution of (5.1). Standard energy estimates give a first control of the solution of the form

$$|u(t_*,r,\omega)|\lesssim \mathrm{e}^{Ct_*}.$$

It is then easy to see that the Fourier transform of u

$$\hat{u}(\sigma)=\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}\sigma t_*}u(t_*)\,\mathrm{d}t_*$$

is holomorphic in $\Im\sigma>C$. We now take Fourier transform of Equation (5.1) to obtain:

$$\hat{\Box}_g(\sigma)\hat{u}(\sigma)=\hat{f}(\sigma).$$

We then can write

$$u(t_*,x)=\frac{1}{2\pi}\int_{\Im\sigma=C+1}\mathrm{e}^{-\mathrm{i}\sigma t_*}\hat{\Box}_g^{-1}(\sigma)\hat{f}(\sigma)\,\mathrm{d}\sigma.$$

This doesn't give so far any better behavior of u, but $\hat{\Box}_g^{-1}(\sigma)$ is a meromorphic family of bounded operators and we can shift the contour to $\Im\sigma=-\delta$. The residue theorem then gives the expansion in (5.3). The most delicate part is the estimate of the error, which requires in particular high frequency estimates of the resolvent $\hat{\Box}_g^{-1}(\sigma)$.

Remark 5.4. By choosing s large enough the strip where the operators form an analytic Fredholm family can be made arbitrary large. For $\Im\sigma=-C$ and C large enough this line lies below an infinite number of resonances. The sum in (5.3) is then infinite, but still converges in some appropriate sense, see [19].

5.2. *Fredholm operators*

Let us recall that a Fredholm operator $T:X\to Y$ is a bounded operator with closed rank such that kernel and cokernel coker $=Y/\mathrm{range}(T)$ are finite dimensional. The index is then defined as $\mathrm{ind}\,T=\dim\ker T-\dim\mathrm{coker}\,T$. The situation is particularly interesting when we have an analytic family $D\ni z\mapsto T(z)$ of Fredholm operators. The index is then a continuous function of z and the Fredholm alternative says that either $T(z)$ is invertible for no $z\in D$ or $F^{-1}(z)$ form a meromorphic

family of operators. Typically elliptic operators are Fredholm between suitable Sobolev spaces and vice versa the Fredholm property between these Sobolev spaces requires ellipticity. We refer to [76, Section 19.2] for details.

The operators $\hat{P}(\sigma)$ we are interested in are not elliptic. An analytic Fredholm theory can nevertheless be developed as observed by Vasy [23]. A central observation is that the Fredholm theory follows from suitable estimates. A general setting is the following, see e.g. [44, Lemma 6.23]:

Lemma 5.5. *Let $X_0 \subset X_1 \subset X_2$ and $Y_0 \subset Y_1 \subset Y_2$ be Banach spaces (with continuous dense inclusions). Let $P : X_1 \to Y_2$ be a bounded operator such that $P|_{X_0}$ is bounded from X_0 to Y_1. We assume that both inclusions $X_1 \subset X_2$ and $Y_0 \subset Y_1$ are compact and that there exists $C > 0$ such that for all $u \in X_1$ and all $v \in Y_1^*$:*

$$\begin{aligned} \|u\|_{X_1} &\le C(\|Pu\|_{Y_1} + \|u\|_{X_2}), \\ \|v\|_{Y_1^*} &\le C(\|P^* v\|_{X_1^*} + \|v\|_{Y_0^*}). \end{aligned}$$

Under these assumptions, P is Fredholm as an operator between the Banach space $\mathscr{X} := \{u \in X_1; Pu \in Y_1\}$ (endowed with the norm $\|u\|_{\mathscr{X}}^2 = \|u\|_{X_1}^2 + \|Pu\|_{Y_1}^2$) and Y_1.

Let $\beta = \min\{(1/\kappa_-), (1/\kappa_+)\}$ if $\Im\sigma \ge 0$, $\beta = \max\{(1/\kappa_-), (1/\kappa_+)\}$ if $\Im\sigma \le 0$. We will obtain the following global estimates.

(1) Let $s > s_0 > (1/2) - \beta\Im\sigma$. Then we have

$$\|u\|_{\bar{H}^s(X)} \le C(\|\hat{\Box}_g(\sigma)u\|_{\bar{H}^{s-1}(X)} + \|u\|_{\bar{H}^{s_0}(X)}). \tag{5.4}$$

(2) Let $s_0 < s' < (1/2) + \beta\Im\sigma$. Then we have:

$$\|u\|_{\dot{H}^{s'}(X)} \le C(\|\hat{\Box}_g^*(\sigma)u\|_{\dot{H}^{s-1}(X)} + \|u\|_{\dot{H}^{s_0}(X)}). \tag{5.5}$$

The compact Sobolev embedding $\bar{H}^s(X) \hookrightarrow \bar{H}^{s_0}(X)$ and Lemma 5.5 then imply the Fredholm property. In the following we will only show (5.4), the arguments for (5.5) are quite similar.

5.3. *Bounded frequency analysis*

Let us recall that we have near $r = r_\pm$:

$$\hat{\Box}_g(\sigma) = r^{-2}D_r r^2 F D_r \pm 2\sigma r^{-2} D_r r^2 - r^{-2}\Delta_{\mathbb{S}^2}.$$

We will apply the unitary transform $U : L^2((r_-, r_+)\times\mathbb{S}^2, r^2\,\mathrm{d}r\,\mathrm{d}\omega) \to L^2((r_-, r_+)\times\mathbb{S}^2, \mathrm{d}r\,\mathrm{d}\omega)$, $u \mapsto ru$ and then work with the operator

$$r\hat{\Box}_g(\sigma)r^{-1} = r^{-1}D_r r^2 F D_r r^{-1} \pm 2\sigma r^{-1} D_r r - r^{-2}\Delta_{\mathbb{S}^2},$$

which we call again $\hat{\Box}_g(\sigma)$ in the following. The above operator has principal classical symbol (ξ is now dual to r):

$$\sigma_2(\hat{\Box}_g(\sigma)) = F\xi^2 + r^{-2}|\eta|^2 =: p.$$

The operator is therefore elliptic over all points (r, ω) of the base manifold with $r_- < r < r_+$. We will distinguish the elliptic and non elliptic regions.

5.3.1. *Elliptic type estimates*

Proposition 5.6. *Let $\chi \in C_0^\infty(r_- + \delta, r_+ - \delta)$ for some $\delta > 0$. Then we have*

$$\|\chi u\|_{\bar{H}^s} \lesssim \|\hat{\Box}_g(\sigma)u\|_{\bar{H}^{s-2}} + \|u\|_{\bar{H}^{s-1}}.$$

Proof. Let $\phi \in C_0^\infty((-1,1))$ with $\phi = 1$ in a neighborhood of 0. We write

$$\chi(x) = \chi(x)\phi(p) + \chi(x)(1-\phi(p)).$$

Note that $\chi(r)\phi(p) \in S^{-\infty}$, $\chi(x)(1-\phi(p)) \in S^0$. By pseudodifferential calculus we only have to consider the second term. Now,

$$Op(\chi(x)(1-\phi(p))) = Op\left(\chi(x)\frac{1-\phi(p)}{p}\right)\hat{\Box}_g(\sigma) + R,\ R \in \Psi^{-1}.$$

The estimate for $s = 0$ then follows by pseudodifferential calculus. Higher order estimates are obtained by applying $Op(\xi^2 + |\eta|^2)$. □

5.3.2. *Radial point estimates*

The characteristic manifold is

$$\Sigma = \{(r,\omega,\xi,\eta) \in T^*X \setminus 0; p(r,\omega,\xi,\eta) = 0\}$$

and it splits into positive and negative energy shells:

$$\Sigma = \Sigma_+ \cup \Sigma_-, \Sigma_\pm = \Sigma \cap \{\pm\xi > 0\}.$$

As $F(r_-) = 0$ there is a lack of ellipticity at $r = r_-$. Note also that we have

$$H_p = (-F'\xi^2 + 2r^{-3}|\eta|^2)\partial_\xi + 2\xi F\partial_r + r^{-2}H_{|\eta|^2}.$$

The

$$\Lambda_\pm^\pm = \left\{(r_\pm,\omega,\xi,0); \xi \in \mathbb{R}^*, \omega \in \mathbb{S}^2\right\} \cap \Sigma_\pm$$

are the radial points of H_p. The lower index indicates the point on the base manifold, the upper index the energy level. It is not difficult to see that Λ_-^+ and Λ_+^- are sources of the hamiltonian vector field whereas Λ_-^- and Λ_+^+ are sinks.

Proposition 5.7. *Suppose that $u \in \bar{H}^s$ with $s > (1/2) - (1/\kappa_\pm)\Im\sigma$. Let $\chi_\pm \in C_0^\infty((r_\pm - \delta, r_\pm + \delta))$. Then for $\delta > 0$ sufficiently small we have:*

$$\|\chi_\pm u\|_{\bar{H}^s} \lesssim \|\hat{\Box}_g(\sigma)u\|_{\bar{H}^{s-1}} + \|u\|_{\bar{H}^{s-\frac{1}{2}}}.$$

Remark 5.8.
(1) The condition $s > (1/2) - (1/\kappa_\pm)\Im\sigma$ gives the condition (5.2) in Theorem 5.1.
(2) Let us try to understand where the condition $s > (1/2) - \beta\Im\sigma$ comes from. As a toy model we consider

$$P_\sigma = D_x x D_x - \sigma D_x, \quad x \in (-1,1).$$

In this case $\beta = 1$. We have

$$P_\sigma(x \pm \mathrm{i}0)^{\mathrm{i}\sigma} = 0.$$

We have $(x \pm \mathrm{i}0)^{\mathrm{i}\sigma} \in H^{(1/2-\Im\sigma)-}$ but not in H^s for $s > (1/2) - \Im\sigma$. Imposing Sobolev regularity $s > (1/2) - \Im\sigma$ thus excludes these solutions.
(3) The estimate is a very simplified version of the radial point estimate obtained by Vasy, see [23, Proposition 2.3]. This kind of estimates goes back to Melrose, see [81]. They are central in Vasy's method concerning the Fredholm theory for non elliptic problems. We also refer to Zworski [82] for a review of this method.

Main ideas of the proof. We will prove the proposition only for a model problem. We will consider the operator which is given near $r = r_\pm$ by

$$\hat{P}(\sigma) = r^{-1}D_r r^2 F D_r r^{-1} \pm 2\sigma r^{-1} D_r r$$

and $\hat{P}(\sigma)$ acts on $L^2((r_- - \epsilon, r_+ + \epsilon); dr)$.

Also to avoid technical issues we will suppose $u \in C_0^\infty((r_- - \epsilon, r_+ + \epsilon))$. We prove the proposition at $r = r_-$, the proof at $r = r_+$ being essentially the same. We also put $x = r - r_-$. Let χ_- be as in

the hypotheses with $x\chi'_-(x) \le 0$. We drop the index—in the following. Let $\psi \in C^\infty(\mathbb{R})$, $\psi(x) = 0$ for $x \le 0$ and $\psi(x) = 1$ for $x \ge 1$.

$$A = C^* C, \quad C = Op(c), \quad c := \xi^{s-\frac{1}{2}}\psi(\xi)\chi(x).$$

Let $a := \sigma_{2s-1}(A)$. Now,

$$\begin{aligned} 2\Im\langle \hat{P}(\sigma)u, Au\rangle &= \langle \mathrm{i}(\hat{P}^*(\sigma)A - A\hat{P}(\sigma))u, u\rangle \\ &= \langle \mathrm{i}[\hat{P}(\sigma), A]u, u\rangle + \langle \mathrm{i}(\hat{P}^*(\sigma) - \hat{P}(\sigma))Au, u\rangle \\ &= \langle \mathrm{i}[\hat{P}(\sigma), A]u, u\rangle - 4\Im\sigma\langle r^{-1}D_x rAu, u\rangle. \end{aligned} \tag{5.6}$$

The symbolic calculation gives

$$H_p a = -(2s-1)\xi^{2s}\psi^2(\xi)F'\chi^2(x) - 2F'\psi(\xi)\psi'(\xi)\xi^{2s+1}\chi^2(x) + 4\xi^{2s}\psi^2(\xi)F\chi(x)\chi'(x). \tag{5.7}$$

Note that the second term is in $S^{-\infty}$ and that the third term is negative.

We have

$$\langle \mathrm{i}[\hat{P}(\sigma), A]u, u\rangle = \langle Op(H_p a)u, u\rangle + \langle Ru, u\rangle, \quad R \in \Psi^{2s-1}.$$

Applying the sharp Gårding inequality, Theorem 4.4, we find

$$\langle \mathrm{i}[\hat{P}(\sigma), A]u, u\rangle - 4\Im\sigma\langle r^{-1}D_x rAu, u\rangle \le (-(2s-1)2\kappa_- + \tilde{\epsilon}(\delta) - 4\Im\sigma)\|Op(\psi(\xi)\chi(x))u\|^2_{\bar{H}^s} + \|u\|^2_{\bar{H}^{s-\frac{1}{2}}},$$

where $\tilde{\epsilon}(\delta) \to 0$, $\delta \to 0$. Here and in the following $\bar{H}^s = \bar{H}^s((r_- - \epsilon, r_- + \epsilon))$.

We also have

$$|2\Im\sigma\langle \hat{P}(\sigma)u, Au\rangle| \le C_{\hat{\epsilon}}\|\hat{P}(\sigma)u\|^2_{\bar{H}^{s-1}} + \hat{\epsilon}\|Op(\psi(\xi)\chi(x))u\|^2_{\bar{H}^s} + C\|u\|^2_{\bar{H}^{s-1}}.$$

Putting everything together and choosing $\delta, \hat{\epsilon}$ small enough gives:

$$\|Op(\psi(\xi)\chi(x))u\|^2_{\bar{H}^s} \le C\left(\|\hat{P}(\sigma)u\|^2_{\bar{H}^s} + \|u\|^2_{\bar{H}^{s-\frac{1}{2}}}\right). \tag{5.8}$$

We now want to obtain the same estimate with $Op(\psi(\xi)\chi(x))$ replaced by $\chi(x)$. To show the estimate for $Op(\psi(-\xi)\chi(x))$ we can replace $a(x,\xi)$ by $a(x,-\xi)$, then the above proof works for $-\hat{P}(\sigma)$ instead of $\hat{P}(\sigma)$. Writing $1 = \psi(\xi) + \psi(-\xi) + \psi_0(\xi)$ and using that $\psi_0(\xi) \in S^{-\infty}$ we see that we can indeed replace $Op(\psi(\xi)\chi(x))$ by $\chi(x)$. □

Remark 5.9. *Using a regularization procedure one can show that it is sufficient to suppose $Cu \in L^2$. What is usually called a radial point estimate is rather (5.8), the important property of $B = Op(\psi(\xi)\chi(x))$ is that backward bicharacteristics starting in $WF'(B)$ tend to Λ^+_-, where C has to be elliptic. In this very simplified model this is used in the proof via the sign of $4\xi^{2s}\psi^2(\xi)F\chi(x)\chi'(x)$ in (5.7). Also $\hat{P}(\sigma)$ can be replaced by $G\hat{P}(\sigma)$ with $G \in \Psi^0$ elliptic at Λ^+_- as long as the closure of all backward bicharacteristics starting in $WF'(B)$ is contained in the elliptic set of G. We refer to [23] for details.*

5.3.3. *Proof of Theorem 5.1*

Putting the estimates in Sections 5.3.1 and 5.3.2 together we obtain the following global estimates:

Proposition 5.10. *Let $u \in \bar{H}^s$ and $s > 1/2 - \beta\Im\sigma$. Then we have*

$$\|u\|_{\bar{H}^s(X)} \le C\left(\|\hat{\Box}_g(\sigma)u\|_{\bar{H}^{s-1}(X)} + \|u\|_{\bar{H}^{s-\frac{1}{2}}(X)}\right).$$

Proof. Let us first choose $\chi_\pm$ as in Proposition 5.7 and $\chi_\pm = 1$ on $(r_\pm - (\delta/2), r_\pm + (\delta/2))$. We then choose $\hat{\chi}_\pm$ with $\mathrm{supp}\hat{\chi}_- \subset [r_- - \epsilon, r_- - (\delta/2))$, $\mathrm{supp}\hat{\chi}_+ \subset (r_+ + (\delta/2), r_+ + \epsilon]$ and $\chi_0 \in C_0^\infty(r_- + (\delta/2), r_+ - (\delta/2))$ with $\hat{\chi}_- + \chi_- + \chi_0 + \chi_+ + \hat{\chi}_+ = 1$. We then use Proposition 5.6 on $\mathrm{supp}\chi_0$ and Proposition 5.7 on $\sup\chi_\pm$. Concerning $\mathrm{supp}\hat{\chi}_\pm$ let us observe that the symbol $p = F\xi^2 + r^{-2}|\eta|^2$ is hyperbolic there, because F is negative on this interval. The estimate follows then from classical hyperbolic estimates, we refer to [44] for details. □

Proof of Theorem 5.1. We obtain in a similar way a suitable estimate also for the adjoint operator. The two estimates then give the Fredholm property. To show that the operators are invertible for $\Im\sigma > 0$ we have to show that kernel and cokernel are empty. We only consider the kernel, the proof for the cokernel is similar. Let $\Im\sigma > 0$, $u \in \bar{H}^s$ and $\hat{\Box}_g(\sigma)u = 0$. Then

$$F\tilde{\Box}_g(\sigma)\mathrm{e}^{-\mathrm{i}\sigma h(r)r_*(r)}u(r,\omega) = 0,$$

where h is some smooth function with $h(r) = 1$ for $r \le 3\mathfrak{m}$ and $h(r) = -1$ for $r \ge 4\mathfrak{m}$ and $\tilde{\Box}_g(\sigma)$ denotes the Fourier transform with respect to t:

$$F\tilde{\Box}_g(\sigma) = r^{-2}FD_r(r^2F)D_r - r^{-2}F\Delta_{\mathbb{S}^2} - \sigma^2.$$

$F\tilde{\Box}_g(0)$ is a selfadjoint positive operator on $L^2((r_-, r_+)\times\mathbb{S}^2; F^{-1}r^2\,\mathrm{d}r\,\mathrm{d}\omega^2)$. But for $\Im\sigma > 0$ we have $\mathrm{e}^{-\mathrm{i}\sigma h(r)r_*(r)}u(r,\omega) \in L^2((r_-, r_+)\times\mathbb{S}^2; F^{-1}r^2\,\mathrm{d}r\,\mathrm{d}\omega^2)$, which is a contradiction. Thus $u|_{(r_-,r_+)\times\mathbb{S}^2} = 0$. To see that u really is zero on the whole manifold $(r_- - \epsilon, r_+ + \epsilon)\times\mathbb{S}^2$ we can follow [82, Lemma 1], see [43, Proof of Theorem 4.5] for details. The index 0 property then follows from the continuous dependence of the index on σ. □

5.4. *High frequency analysis: trapping*

Whereas the bounded frequency analysis is sufficient to prove Theorem 5.1, the proof of Theorem 5.2 requires a precise control of the estimates when $\Re\sigma \to \infty$. This requires a precise analysis of the semiclassical flow whose projection on the base manifold gives the geodesic flow. The analysis then becomes more complicated, but the general structure of elliptic, propagation and radial point estimates remains similar. However there is a new aspect which we want to discuss in this subsection which is trapping. To analyze the trapping it is easier to go back to the initial Boyer Lindquist coordinates. We will present a toy model, the proof for the existence of a resonance free strip follows very closely the proof of Dyatlov [26] for the general situation.

5.4.1. *Resonance free strip*

Fourier transforming with respect to t and fixing the angular momentum ℓ leads to considering the operator:

$$\tilde{P}_\ell - \sigma^2 = D_{r_*}^2 + V\ell(\ell+1) - \sigma^2, \quad V = \frac{F}{r^2}. \tag{5.9}$$

We introduce the semiclassical parameter $h^2 = 1/\ell(\ell+1)$. As explained in Section 2.4.3 trapping is concentrated at $r_* = 0$. We introduce a Taylor development of V at $r_* = 0$. We will put $x = r_*$ in the following. This leads to considering the local model

$$\tilde{P} - \lambda = h^2D_x^2 - x^2 - \lambda, \quad \lambda = \mathcal{O}(h), \tag{5.10}$$

which we will use as a toy model in the following. Note that in the model (5.9) the critical energies are $\sigma^2 \sim V(3\mathfrak{m})\ell(\ell+1)$, after rescaling this becomes $\lambda \sim \sqrt{V(3\mathfrak{m})}$. In the toy model however the critical energy is zero, we therefore consider $\lambda = \mathcal{O}(h)$. We obtain semiclassical estimates that can be translated into high frequency estimates in the original scaling.

The symbol writes $\xi^2 - x^2 = (\xi - x)(\xi + x)$ which suggests that the operator is unitary equivalent to $P = h(xD_x + D_xx)$. We can indeed introduce the transform

$$Uu(x) = \frac{1}{\sqrt{\pi}h}\mathrm{e}^{-\mathrm{i}\frac{x^2}{2h}}\int \mathrm{e}^{\mathrm{i}\frac{\sqrt{2}xy}{h}}\mathrm{e}^{-\mathrm{i}\frac{y^2}{2h}}u(y)\,\mathrm{d}y.$$

It can be checked that $U : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is an isometry and that $\tilde{P}U = UP$. We will in the following consider the operator $P = (h/2)(xD_x + D_xx)$. Note that

$$\sigma_h(P) = x\xi =: p, \quad H_p = -\xi\partial_\xi + x\partial_x.$$

Thus the Hamiltonian equations can be explicitly solved, $\xi(s) = \mathrm{e}^{-t}\xi_0$, $x(s) = \mathrm{e}^{t}x_0$, where (x_0,ξ_0) are the initial data. In the abstract setting of Section 2.4.4 this means that we have:

$$\Gamma_+ = \{\xi = 0\}, \quad \Gamma_- = \{x = 0\}, \quad K = \{x = \xi = 0\}.$$

To analyze the situation near the trapped set, we will consider the situation on a compact manifold and add an absorption operator. Concretely, we put $X = \mathbb{S}^1 = \mathbb{R}/(6\mathbb{Z})$; we view X as the interval $[-3,3]$ with the endpoints glued together. We add an absorption operator $Q = Op(q)$, $q \in C^\infty(T^*X)$ such that $q(x,\xi) = 0$ for all $|(x,\xi)| \le 1$ and $q(x,\xi) = 1$ for all $|(x,\xi)| \ge 2$. We will now consider the operator $P - \mathrm{i}Q - \lambda$ on X. Let $R(\lambda) = (P - \mathrm{i}Q - \lambda)^{-1}$.

Lemma 5.11. *For each $\epsilon > 0$ and h small enough depending on ϵ we have*

$$\|R(\lambda)\|_{L^2\to L^2} = o(h^{-2}) \quad \text{if } |\lambda| = \mathcal{O}(h), \quad \Im\lambda > -(\tfrac{1}{2} - \epsilon)h. \tag{5.11}$$

Remark 5.12. *Let us note that the same bound on the resolvent holds in general situations of normally hyperbolic trapping as described in Section 2.4.4, see [26]. We will give here the same proof for the model operator, which has the advantage of avoiding the technical issues while keeping the main ideas of the proof. Further gaps can be obtained for r-normally hyperbolic trapping, see also [26].*

To prove the lemma we will proceed by contradiction. Suppose that (5.11) does not hold. Then for some $\epsilon > 0$ small enough, there exist sequences $h_j, h_j \to 0$, $\lambda(h_j)$, $u = u(h_j) \in L^2$ with the following properties

$$\|u\|_{L^2} = 1, \quad |\lambda| = \mathcal{O}(h), \quad \Im\lambda > -(\tfrac{1}{2} - \epsilon)h \tag{5.12}$$

and

$$(P - \mathrm{i}Q - \lambda)u = \mathcal{O}(h^2). \tag{5.13}$$

The interesting properties are obtained when commuting $(P-\lambda)u$ with the defining functions of $\Gamma_\pm$. Indeed we have

$$hD_x(P-\lambda)u = (P - \mathrm{i}h - \lambda)hD_x u.$$

Now suppose that $\Im\lambda \ge -((1/2) - \epsilon)h$ for some $\epsilon > 0$. Then $h + \Im\lambda \ge ((1/2) + \epsilon)h$ and we can write

$$hD_x u = (P - \mathrm{i}h - \lambda)^{-1} hD_x(P-\lambda)u. \tag{5.14}$$

A simple application of the spectral theorem gives

$$\|hD_x u\| \lesssim h^{-1}\|hD_x(P-\lambda)u\|. \tag{5.15}$$

Note that the corresponding commutation relation with x

$$x(P-\lambda) = (P + \mathrm{i}h - \lambda)xu.$$

doesn't give an analogous estimate because of the different sign.

The estimate (5.15) can be microlocalized. Let for $\delta \ll 1$

$$U_\delta = \{|\xi| < \delta, |x| < \delta\}.$$

Note that

$$(P-\lambda)u = \mathcal{O}(h^2) \text{ microlocally on } U_{2\delta}. \tag{5.16}$$

Let $\phi \in C_0^\infty((-4\delta,4\delta)), \phi = 1$ on $[-3\delta,3\delta]$, $\theta_+ = \xi\phi(x)\phi(\xi)$, $\Theta_+ = Op_h(\theta_+)$. Applying Θ_+ to (5.16) and using that $[\Theta_+, P] = \mathrm{i}hOp_h(H_p\theta_+) + \mathcal{O}(h^2) = -\mathrm{i}h\Theta_+ + \mathcal{O}(h^2)$ microlocally on $U_{2\delta}$ together with (5.16) gives an appropriate microlocalized version of (5.15):

$$\Theta_+ u = \mathcal{O}(h), \quad \text{microlocally on } U_{2\delta}. \tag{5.17}$$

A central lemma is then the following:

Lemma 5.13. *There exists a constant C such that for each $\delta_0 > 0$,*

$$\mu(U_\delta \cap \{|x| < \delta_0\}) \le C\delta_0. \tag{5.18}$$

Proof. We first note that for $a \in C_0^\infty(\Gamma_+ \cap U_\delta)$ we have:

$$\left| \int_{\Gamma_+ \cap U_\delta} H_{\theta_+} a \, \mathrm{d}\mu \right| \le C\|a\|_\infty. \tag{5.19}$$

This follows indeed from Lemma 4.9 and (5.17). Now choose $a = \tilde{a}(x)\chi(\xi)$, with $\chi \in C_0^\infty((-\delta,\delta))$, $\chi(0) = 1$ and

$$\operatorname{supp}\tilde{a} \subset (-2\delta_0, \delta_1), \quad \|\tilde{a}\|_\infty \le 1, \quad \partial_x \tilde{a} \ge -\frac{2}{\delta_1}, \quad \partial_x \tilde{a} \ge \frac{1}{3\delta_0} \quad \text{for } |x| \le \delta_0$$

for some fixed $\delta_1 \in (0,\delta)$ independent of δ_0. Noting that $\theta_+ = \xi$ on U_δ we find

$$\int_{\Gamma_+ \cap U_\delta} H_{\theta_+} a \, \mathrm{d}\mu \ge \frac{1}{3\delta_0}\mu(\{|x| \le \delta_0\}) - \frac{2}{\delta_1}\mu(\{\delta_0 \le |x| \le \delta_1\}).$$

The LHS is bounded by a δ_0-independent constant by (5.19); so is the second term on the RHS (since μ is a finite measure). Multiplying both sides by $3\delta_0$ we obtain (5.18). □

Lemma 5.14.

$$\mu(U_\delta \setminus \Gamma_+) = 0, \quad \mu(U_\delta) > 0.$$

Proof. The lemma follows from two type of estimates

(1) Take $A \in \Psi_h^0(\mathbb{R})$ such that $WF_h'(A) \cap \{p = 0\} \cap \Gamma_+ = \emptyset$. Then

$$\|Au\|_{L^2} \le Ch^{-1}\|(P - \mathrm{i}Q - \lambda)u\|_{L^2} + \mathcal{O}(h^\infty)\|u\|_{L^2}.$$

This follows from the elliptic estimate [83, Proposition 2.4] and propagation of singularities with a complex absorbing operator [83, Proposition 2.4], since for each $(x,\xi) \in WF_h'(A)$, there exists $t \ge 0$ such that $\mathrm{e}^{-tH_p}(x,\xi) \in \{p - \mathrm{i}q \ne 0\}$.

(2) Let $B \in \Psi_h^{\mathrm{comp}}(X)$ such that $\sigma_h(B) \ne 0$ on $\{p = 0\} \cap K$:

$$\|u\|_{L^2} \le Ch^{-1}\|(P - \mathrm{i}Q - \lambda)u\|_{L^2} + \|Bu\|_{L^2}.$$

To prove this second estimate we use the elliptic estimate and propagation of singularities, noting that for each $(x,\xi) \in T^*\mathbb{R}$, there exists $t \ge 0$ such that $\mathrm{e}^{-tH_p}(x,\xi) \in \{p - \mathrm{i}q \ne 0\} \cup \{\sigma_h(B) \ne 0\}$, see [41, Lemma 4.1].

□

Proof of Lemma 5.11. Starting with (5.13) and passing to a subsequence, we may assume that

$$u \to \mu, \quad h^{-1}\Im\lambda \to \omega \in \mathscr{C}.$$

By Lemma 5.14 we know that

$$\mu(U_\delta) > 0, \quad \mu(U_\delta \setminus \Gamma_+) = 0.$$

We can now apply Lemma 4.9 to see that for each $a \in C_0^\infty(U_\delta)$ we have

$$\int_{\Gamma_+ \cap U_\delta} H_p a \, \mathrm{d}\mu = 2\Im\omega \int_{\Gamma_+ \cap U_\delta} a \, \mathrm{d}\mu.$$

For $t \ge 0$, since $\mathrm{e}^{-tH_p}(\Gamma_+ \cap U_\delta) \subset \Gamma_+ \cap U_\delta$, we have

$$\mu(\mathrm{e}^{-tH_p}(U_\delta)) = \mathrm{e}^{2t\Im\omega}\mu(U_\delta).$$

By Lemma 5.13, as $t \to \infty$

$$\mu(\mathrm{e}^{-tH_p}(U_\delta)) \le \mu(U_\delta \cap \{|x| < \delta\mathrm{e}^{-t}\}) \le C\mathrm{e}^{-t}.$$

However, since $\Im\omega \ge -(1/2) + \epsilon$, we see that $\mathrm{e}^{2t\Im\omega}$ decays exponentially slower than e^{-t}, as $t \to \infty$. Since $\mu(U_\delta) > 0$ we arrive at a contradiction, finishing the proof of the Proposition. □

5.4.2. *Approximation of resonances generated by the trapping*

As we have just seen the relatively mild trapping leads to a resonance free strip. Nevertheless the trapping creates resonances that we now want to compute in a high frequency limit. The resonances can be computed as the eigenvalues of a scaled operator which is no longer selfadjoint. The operator P coming from the De Sitter Schwarzschild metric is an operator with holomorphic coefficients

$$P(w, hD_w; h) = (hD_w)^2 + V(w, h); \quad V(w, h) = Fr^{-2}(1 + h^2 r\partial_r F). \tag{5.20}$$

Here w replaces the Regge–Wheeler coordinate r_*. We put $\Gamma_\theta = \mathrm{e}^{\mathrm{i}\theta}\mathbb{R}$ and we parametrize Γ_θ by x, $w = \mathrm{e}^{\mathrm{i}\theta} x$. The scaled operator is then defined by $P_\theta(x, hD_x; h) = P(w, hD_w; h)|_{\Gamma_\theta}$. Note that the RHS of this equation is well defined because all the coefficients are holomorphic. The important proposition is the following (see [18]):

Proposition 5.15. *For $-2\theta < arg\, z < 2\pi - 2\theta$ the operator*

$$P_\theta(x, hD_x; h) - z : H^2(\Gamma_\theta) \to L^2(\Gamma_\theta),$$

is a Fredholm operator of index 0. *The eigenvalues of $P_\theta(x, hD_x; h)$ in $-2\theta < \arg z < 2\pi - 2\theta$ coincide with the poles of the meromorphic continuation of $(P_\theta(x, hD_x; h) - z)^{-1}$ there.*

To figure out what is a good choice for θ we come back to our toy model $\tilde{P} = h^2 D_x^2 - x^2$ in the previous section. We then see that a good choice is $\theta = \pi/4$. Indeed we then find $P_\theta = -\mathrm{i}(h^2 D_x^2 + x^2)$, which is just $-\mathrm{i}$ times the (semiclassical) harmonic oscillator and its eigenvalues are of course known. Putting all this together leads to the following characterization of the high frequency resonances in the De Sitter Schwarzschild case (see [18]):

Proposition 5.16. *There exists $K > 0$ and $\theta > 0$ such that for any $C > 0$ there exists an injective map, $\tilde{b}$, from the sets of pseudo-poles*

$$\frac{(1 - 9\Lambda\mathfrak{m}^2)^{1/2}}{3^{\frac{3}{2}}\mathfrak{m}} \left(\pm\mathbb{N} \pm \frac{1}{2} - \mathrm{i}\frac{1}{2}\left(\mathbb{N}_0 + \frac{1}{2}\right)\right)$$

into the set of poles of the meromorphic continuation of $(P - \sigma^2)^{-1}$ such that the poles in

$$\Omega_C = \{\sigma : \Im\sigma > -C, |\sigma| > K, \Im\sigma > -\theta|\Re\sigma|\}$$

are in the image of $\tilde{b}$ and for $\tilde{b}(\nu) \in \Omega_C$,

$$\tilde{b}(\nu) - \nu \to 0, \quad when\ |\nu| \to \infty.$$

If $\Re\nu = ((1 - 9\Lambda\mathfrak{m}^2)^{1/2}/3^{3/2}\mathfrak{m})(\pm\ell \pm (1/2))$, $\ell = 1, 2, \ldots$, then the corresponding pole, $\tilde{b}(\nu)$ has multiplicity $2\ell + 1$.

5.5. *More general situations*

5.5.1. *De Sitter Kerr metric*

The proof for the equivalent of Theorems 5.1 and 5.2 for the De Sitter Kerr metric follows the same lines as in the De Sitter Schwarzschild case. The radial point estimates are now formulated in terms of propagation estimates which permit to propagate regularity out of these points. As already mentioned the trapping keeps the same structure and the proof for the high energy resolvent estimates follows the same lines. At first glance the new aspect seems to be superradiance. Nevertheless thanks to the very robust Fredholm setting this point is now reduced to the analysis of mode solutions.

5.5.2. *Kerr metric*

The Kerr case is asymptotically Minkowskian and one has to deal with this aspect. As an immediate consequence, the underlying resolvent doesn't possess a meromorphic extension in a strip beyond the real axis. It has nevertheless good C^k properties down to the real axis, see [33,34] and [39]. Another important aspect is that we find a radial point structure of the Hamiltonian flow also at infinity.

5.5.3. *Other fields*

As long as the fields are massless and not charged, most of the here presented aspects remain valid also for other hyperbolic field equations than the wave equation. Note that the presence of mass or charge of the field can also create growing modes, we refer to Shlapentokh-Rothmann for the Klein Gordon equation on the Kerr metric [84] and to Besset–Häfner [85] for the charged Klein Gordon equation on the De Sitter Reissner–Nordström or De Sitter Kerr Newman black hole. In the case of linearized gravity, one has of course to deal with the gauge. The presented methods are well adapted to generalized wave map gauges. In these generalized wave map gauges the Fourier linearized Einstein operator $\widehat{L}_g(\sigma)$ will have a kernel for $\sigma = 0$ consisting of linearized (De Sitter) Kerr solutions and some special gauge solutions. The existence of this kernel makes the analysis of $\widehat{L}_g^{-1}(\sigma)$ more delicate close to $\sigma = 0$.

5.5.4. *Perturbations of the (De Sitter) Kerr metric*

As long as the metric remains stationary and the important structure of radial points, trapping etc. is not changed the analysis remains more or less the same on the linear level. Note that in contrast to the trapping the radial point structure is a priori not stable with respect to perturbations; however the estimates are, see [23, Remark 2.5] for a discussion of this point. A delicate part is the analysis of the mode solutions which most often is analyzed on the exact spacetimes by separation of variables and ode techniques. This can however often be analyzed in a perturbative regime.

If the spacetime itself is not stationary but say decays to a (De Sitter) Kerr solution, the here presented analysis has to be completed by a second step. Note however that this second step can be purely microlocal, see [63]. These aspects go beyond the scope of this review.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

Acknowledgements

The author thanks Peter Hintz and Pascal Millet for helpful discussions on this review. He also acknowledges support from the ANR funding ANR-20-CE40-0018-01.

References

[1] J. J. Duistermaat, L. Hörmander, "Fourier integral operators. II", *Acta Math.* **128** (1972), p. 183-269.

[2] M. J. Radzikowski, "Micro-local approach to the Hadamard condition in quantum field theory on curved space-time", *Commun. Math. Phys.* **179** (1996), no. 3, p. 529-553.

[3] C. Gérard, M. Wrochna, "Construction of Hadamard states by pseudo-differential calculus", *Commun. Math. Phys.* **325** (2014), no. 2, p. 713-755.

[4] J. Dimock, B. S. Kay, "Classical and quantum scattering theory for linear scalar fields on the Schwarzschild metric. I", *Ann. Phys.* **175** (1987), p. 366-426.

[5] A. Bachelot, "Asymptotic completeness for the Klein–Gordon equation on the Schwarzschild metric", *Ann. Inst. Henri Poincaré, Phys. Théor.* **61** (1994), no. 4, p. 411-441.

[6] A. Bachelot, "The Hawking effect", *Ann. Inst. Henri Poincaré, Phys. Théor.* **70** (1999), no. 1, p. 41-99.

[7] C. Dappiaggi, V. Moretti, N. Pinamonti, "Rigorous construction and Hadamard property of the Unruh state in Schwarzschild spacetime", *Adv. Theor. Math. Phys.* **15** (2011), no. 2, p. 355-447.

[8] D. Häfner, *Creation of Fermions by Rotating Charged Black Holes*, Mém. Soc. Math. Fr., Nouv. Sér., vol. 117, Société Mathématique de France (SMF), Paris, 2009.

[9] C. Gérard, D. Häfner, M. Wrochna, "The Unruh state for massless fermions on Kerr spacetime and its Hadamard property", *Ann. Sci. Éc. Norm. Supér. (4)* **56** (2023), no. 1, p. 127-196.

[10] D. Häfner, J.-P. Nicolas, "Scattering of massless Dirac fields by a Kerr black hole", *Rev. Math. Phys.* **16** (2004), no. 1, p. 29-123.

[11] C. K. M. Klein, "Construction of the Unruh state for a real scalar field on the Kerr–de Sitter spacetime", *Ann. Henri Poincaré* **24** (2023), no. 7, p. 2401-2442.

[12] D. Häfner, "On the scattering theory for a Klein–Gordon equation in the Kerr metric", *Diss. Math.* **421** (2003), p. 1-102 (French).

[13] M. Dafermos, I. Rodnianski, Y. Shlapentokh-Rothman, "A scattering theory for the wave equation on Kerr black hole exteriors", *Ann. Sci. Éc. Norm. Supér. (4)* **51** (2018), no. 2, p. 371-486.

[14] V. Georgescu, C. Gérard, D. Häfner, "Asymptotic completeness for superradiant Klein–Gordon equations and applications to the de Sitter–Kerr metric", *J. Eur. Math. Soc. (JEMS)* **19** (2017), no. 8, p. 2371-2444.

[15] P. Hintz, A. Vasy, "Analysis of linear waves near the Cauchy horizon of cosmological black holes", *J. Math. Phys.* **58** (2017), no. 8, article no. 081509.

[16] D. Christodoulou, S. Klainerman, *The Global Nonlinear Stability of the Minkowski Space*, Princeton Mathematical Series, vol. 41, Princeton University Press, Princeton, NJ, 1993.

[17] A. Bachelot, A. Motet-Bachelot, "Resonances of a Schwarzschild black hole", *Ann. Inst. Henri Poincaré, Phys. Théor.* **59** (1993), no. 1, p. 3-68 (French).

[18] A. S. Barreto, M. Zworski, "Distribution of resonances for spherical black holes", *Math. Res. Lett.* **4** (1997), no. 1, p. 103-121.

[19] J.-F. Bony, D. Häfner, "Decay and non-decay of the local energy for the wave equation on the De Sitter–Schwarzschild metric", *Commun. Math. Phys.* **282** (2008), no. 3, p. 697-719.

[20] R. Melrose, A. S. Barreto, A. Vasy, "Asymptotics of solutions of the wave equation on de Sitter–Schwarzschild space", *Commun. Partial Differ. Equ.* **39** (2014), no. 3, p. 512-529.

[21] S. Dyatlov, "Quasi-normal modes and exponential energy decay for the Kerr–de Sitter black hole", *Commun. Math. Phys.* **306** (2011), no. 1, p. 119-163.

[22] S. Dyatlov, "Exponential energy decay for Kerr–de Sitter black holes beyond event horizons", *Math. Res. Lett.* **18** (2011), no. 5, p. 1023-1035.

[23] A. Vasy, "Microlocal analysis of asymptotically hyperbolic and Kerr–de Sitter spaces (with an appendix by Semyon Dyatlov)", *Invent. Math.* **194** (2013), no. 2, p. 381-513.

[24] A. Vasy, M. Wrochna, "Quantum fields from global propagators on asymptotically Minkowski and extended de Sitter spacetimes", *Ann. Henri Poincaré* **19** (2018), no. 5, p. 1529-1586.

[25] J. Wunsch, M. Zworski, "Resolvent estimates for normally hyperbolic trapped sets", *Ann. Henri Poincaré* **12** (2011), no. 7, p. 1349-1385.

[26] S. Dyatlov, "Spectral gaps for normally hyperbolic trapping", *Ann. Inst. Fourier* **66** (2016), no. 1, p. 55-82.

[27] P. Hintz, "Normally hyperbolic trapping on asymptotically stationary spacetimes", *Probab. Math. Phys.* **2** (2021), no. 1, p. 71-126.

[28] P. Hintz, A. Vasy, "The global non-linear stability of the Kerr–de Sitter family of black holes", *Acta Math.* **220** (2018), no. 1, p. 1-206.

[29] O. Petersen, A. Vasy, "Stationarity and Fredholm theory in subextremal Kerr–de Sitter spacetimes", *SIGMA, Symmetry Integrability Geom. Methods Appl.* **20** (2024), article no. 052.

[30] P. Hintz, "Mode stability and shallow quasinormal modes of Kerr–de Sitter black holes away from extremality", *J. Eur. Math. Soc.* (2024) (Online first).

[31] J.-F. Bony, D. Häfner, "Low frequency resolvent estimates for long range perturbations of the Euclide Laplace", *Math. Res. Lett.* **17** (2010), no. 2, p. 301-306.

[32] J.-F. Bony, D. Häfner, "Local energy decay for several evolution equations on asymptotically Euclidean manifolds", *Ann. Sci. Éc. Norm. Supér. (4)* **45** (2012), no. 2, p. 311-335.

[33] A. Vasy, "Limiting absorption principle on Riemannian scattering (asymptotically conic) spaces, a Lagrangian approach", *Commun. Partial Differ. Equ.* **46** (2021), no. 5, p. 780-822.

[34] A. Vasy, “Resolvent near zero energy on Riemannian scattering (asymptotically conic) spaces, a Lagrangian approach”, *Commun. Partial Differ. Equ.* **46** (2021), no. 5, p. 823-863.
[35] D. Baskin, A. Vasy, J. Wunsch, “Asymptotics of radiation fields in asymptotically Minkowski space”, *Am. J. Math.* **137** (2015), no. 5, p. 1293-1364.
[36] D. Baskin, A. Vasy, J. Wunsch, “Asymptotics of scalar waves on long-range asymptotically Minkowski spaces”, *Adv. Math.* **328** (2018), p. 160-216.
[37] D. Gajic, C. M. Warnick, “Quasinormal modes on Kerr spacetimes”, preprint, 2024, https://arxiv.org/abs/2407.04098.
[38] T. Stucker, “Quasinormal modes for the Kerr black hole”, preprint, 2024, https://arxiv.org/abs/2407.04612.
[39] D. Häfner, P. Hintz, A. Vasy, “Linear stability of slowly rotating Kerr black holes”, *Invent. Math.* **223** (2021), no. 3, p. 1227-1406.
[40] K. S. Thorne, “Disk accretion onto a black hole. 2. Evolution of the hole”, *Astrophys. J.* **191** (1974), p. 507-520.
[41] S. Dyatlov, “Asymptotics of linear waves and resonances with applications to black holes”, *Commun. Math. Phys.* **335** (2015), no. 3, p. 1445-1485.
[42] B. F. Whiting, “Mode stability of the Kerr black hole”, *J. Math. Phys.* **30** (1989), no. 6, p. 1301-1305.
[43] L. Andersson, D. Häfner, B. F. Whiting, “Mode analysis for the linearized Einstein equations on the Kerr metric : the large a case”, *J. Eur. Math. Soc.* (2024) (Online first).
[44] P. Millet, “Optimal decay for solutions of the Teukolsky equation on the Kerr metric for the full subextremal range $|a| < M$”, preprint, 2023, https://arxiv.org/abs/2302.06946.
[45] P. Hintz, “A sharp version of Price’s law for wave decay on asymptotically flat spacetimes”, *Commun. Math. Phys.* **389** (2022), no. 1, p. 491-542.
[46] S. Klainerman, J. Szeftel, “Kerr stability for small angular momentum”, *Pure Appl. Math. Q.* **19** (2023), no. 3, p. 791-1678.
[47] S. Klainerman, J. Szeftel, “Construction of GCM spheres in perturbations of Kerr”, *Ann. PDE* **8** (2022), no. 2, article no. 17.
[48] S. Klainerman, J. Szeftel, “Effective results on uniformization and intrinsic GCM spheres in perturbations of Kerr”, *Ann. PDE* **8** (2022), no. 2, article no. 18.
[49] E. Giorgi, S. Klainerman, J. Szeftel, “Wave equations estimates and the nonlinear stability of slowly rotating Kerr black holes”, preprint, 2022, https://arxiv.org/abs/2205.14808.
[50] D. Shen, “Construction of GCM hypersurfaces in perturbations of Kerr”, *Ann. PDE* **9** (2023), no. 1, article no. 11.
[51] M. Dafermos, G. Holzegel, I. Rodnianski, M. Taylor, “The non-linear stability of the Schwarzschild family of black holes”, preprint, 2021, https://arxiv.org/abs/2104.08222.
[52] O. Gannot, “Quasinormal modes for Schwarzschild-AdS black holes: exponential convergence to the real axis”, *Commun. Math. Phys.* **330** (2014), no. 2, p. 771-799.
[53] G. Holzegel, J. Smulevici, “Quasimodes and a lower bound on the uniform energy decay rate for Kerr-AdS spacetimes”, *Anal. PDE* **7** (2014), no. 5, p. 1057-1090.
[54] M. Dafermos, “The interior of charged black holes and the problem of uniqueness in general relativity”, *Commun. Pure Appl. Math.* **58** (2005), no. 4, p. 445-504.
[55] J. Luk, S.-J. Oh, “Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data. I: The interior of the black hole region”, *Ann. Math. (2)* **190** (2019), no. 1, p. 1-111.
[56] J. Luk, S.-J. Oh, “Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data II: the exterior of the Black Hole region”, *Ann. PDE* **5** (2019), no. 1, article no. 6.
[57] V. Cardoso, J. L. Costa, K. Destounis, P. Hintz, A. Jansen, “Quasinormal modes and strong cosmic censorship”, preprint, 2017, https://arxiv.org/abs/1711.10502.
[58] S. Hollands, R. M. Wald, J. Zahn, “Quantum instability of the Cauchy horizon in Reissner–Nordström–deSitter spacetime”, *Class. Quantum Gravity* **37** (2020), no. 11, article no. 115009.
[59] C. Kehle, Y. Shlapentokh-Rothman, “A scattering theory for linear waves on the interior of Reissner–Nordström black holes”, *Ann. Henri Poincaré* **20** (2019), no. 5, p. 1583-1650.
[60] S. Chandrasekhar, J. B. Hartle, “On crossing the Cauchy horizon of a Reissner–Nordström black-hole”, *Proc. R. Soc. Lond., Ser. A* **384** (1982), no. 1787, p. 301-315.
[61] D. Häfner, M. Mokdad, J.-P. Nicolas, “Scattering theory for Dirac fields inside a Reissner–Nordström-type black hole”, *J. Math. Phys.* **62** (2021), no. 8, article no. 081503.
[62] M. Mokdad, M. Provci, “Scattering of Dirac fields in the interior of Kerr–Newman(-Anti)–de Sitter black holes”, preprint, 2023, https://arxiv.org/abs/2303.03835.
[63] P. Hintz, “Linear waves on asymptotically flat spacetimes. I”, preprint, 2023, https://arxiv.org/abs/2302.14647.
[64] P. Hintz, A. Vasy, “Microlocal analysis near null infinity in asymptotically flat spacetimes”, preprint, 2023, https://arxiv.org/abs/2302.14613.
[65] N. Fenichel, “Persistence and smoothness of invariant manifolds for flows”, *Indiana Univ. Math. J.* **21** (1971), p. 193-226.

[66] M. W. Hirsch, C. C. Pugh, M. Shub, *Invariant Manifolds*, Lecture Notes in Mathematics, vol. 583, Springer, Cham, 1977.

[67] Y. Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires", *Acta Math.* **88** (1952), p. 141-225 (French).

[68] D. Häfner, P. Hintz, A. Vasy, "Correction to: "Linear stability of slowly rotating Kerr black holes"", *Invent. Math.* **236** (2024), no. 1, p. 477-481.

[69] L. Andersson, T. Bäckdahl, P. Blue, S. Ma, "Stability for linearized gravity on the Kerr spacetime", preprint, 2019, https://arxiv.org/abs/1903.03859.

[70] M. Dafermos, G. Holzegel, I. Rodnianski, "The linear stability of the Schwarzschild solution to gravitational perturbations", *Acta Math.* **222** (2019), no. 1, p. 1-214.

[71] P.-K. Hung, J. Keller, M.-T. Wang, "Linear stability of Schwarzschild spacetime: decay of metric coefficients", *J. Differ. Geom.* **116** (2020), no. 3, p. 481-541.

[72] P. Millet, "Geometric background for the Teukolsky equation revisited", *Rev. Math. Phys.* **36** (2024), no. 3, article no. 2430003.

[73] Y. Shlapentokh-Rothman, R. T. da Costa, "Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: frequency space analysis", preprint, 2023, https://arxiv.org/abs/2007.07211.

[74] Y. Shlapentokh-Rothman, R. T. da Costa, "Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: physical space analysis", preprint, 2023, https://arxiv.org/abs/2302.08916.

[75] S. Ma, L. Zhang, "Sharp decay for Teukolsky equation in Kerr spacetimes", *Commun. Math. Phys.* **401** (2023), no. 1, p. 333-434.

[76] L. Hörmander, *The Analysis of Linear Partial Differential Operators. III: Pseudo-Differential Operators*, Grundlehren der mathematischen Wissenschaften, vol. 274, Springer, Cham, 1985.

[77] M. Zworski, *Semiclassical Analysis*, Graduate Studies in Mathematics, vol. 138, American Mathematical Society (AMS), Providence, RI, 2012.

[78] J. Wunsch, "Microlocal analysis and evolution equations: lecture notes from 2008 CMI/ETH summer school April 25, 2013", in *Evolution Equations. Proceedings of the Clay Mathematics Institute Summer School, ETH, Zürich, Switzerland, June 23–July 18, 2008* (D. Ellwood *et al.*, eds.), Clay Mathematics Proceedings, vol. 17, American Mathematical Society (AMS), Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2013, p. 1-72.

[79] P. Hintz, "Introduction to microlocal analysis", Lecture Notes 2023, https://people.math.ethz.ch/~hintzp/notes/micro.pdf.

[80] P. Hintz, *Global analysis and nonlinear wave equations on cosmological spacetimes*, Phd thesis, Stanford university, 2015, https://people.math.ethz.ch/~hintzp/thesis-augmented.pdf.

[81] R. B. Melrose, "Spectral and scattering theory for the Laplacian on asymptotically Euclidean spaces, in Spectral and scattering theory", in *Proceedings of the 30th Taniguchi international workshop, held at Sanda, Hyogo, Japan*, Marcel Dekker, Basel, 1994, p. 85-130.

[82] M. Zworski, "Resonances for asymptotically hyperbolic manifolds: Vasy's method revisited", *J. Spectr. Theory* **6** (2016), no. 4, p. 1087-1114.

[83] S. Dyatlov, M. Zworski, "Dynamical zeta functions for Anosov flows via microlocal analysis", *Ann. Sci. Éc. Norm. Supér. (4)* **49** (2016), no. 3, p. 543-577, http://hdl.handle.net/1721.1/115500.

[84] Y. Shlapentokh-Rothman, "Exponentially growing finite energy solutions for the Klein–Gordon equation on subextremal Kerr spacetimes", *Commun. Math. Phys.* **329** (2014), no. 3, p. 859-891.

[85] N. Besset, D. Häfner, "Existence of exponentially growing finite energy solutions for the charged Klein–Gordon equation on the de Sitter–Kerr–Newman metric", *J. Hyperbolic Differ. Equ.* **18** (2021), no. 2, p. 293-310.

Comptes Rendus
Mécanique

Objet de la revue

Les *Comptes Rendus Mécanique* sont une revue électronique évaluée par les pairs de niveau international, qui couvre l'ensemble des domaines des sciences mécaniques.
Ils publient des articles originaux de recherche, des articles de synthèse, des mises en perspective historiques, des textes à visée pédagogique, ou encore des actes de colloque, en anglais ou en français, sans limite de longueur et dans un format aussi souple que possible (figures, données associées, etc.).
Depuis 2020, les *Comptes Rendus Mécanique* sont publiés avec le centre Mersenne pour l'édition scientifique ouverte, selon une politique vertueuse de libre accès diamant, gratuit pour les auteurs (pas de frais de publication) comme pour les lecteurs (accès libre, immédiat et pérenne).

Directeur de la publication : Étienne Ghys.

Rédacteurs en chef : Jean-Baptiste Leblond, Samuel Forest.

Comité éditorial : Olga Budenkova, Francisco Chinesta, Félix Darve, Florian Gosselin, Éric Lemarchand, Nicolas Moës, Léo Morin, Benoît Perthame, Guillaume Ribert, Géry de Saxcé, Emmanuel Villermaux.

Secrétaire éditoriale : Adenise Lopes.

À propos de la revue

Toutes les informations concernant la revue, y compris le texte des articles publiés qui est en accès libre intégral, figurent sur le site https://comptes-rendus.academie-sciences.fr/mecanique/.

Informations à l'attention des auteurs

Pour toute question relative à la soumission des articles, les auteurs peuvent consulter le site https://comptes-rendus.academie-sciences.fr/mecanique/.

Contact

Académie des sciences
23, quai de Conti, 75006 Paris, France
Tel: (+33) (0)1 44 41 43 72
cr-mecanique@academie-sciences.fr

Pour citer ce numéro thématique : Jérémie Szeftel (éd). Recent advances in general relativity: an issue in memory of Yvonne Choquet-Bruhat. Comptes Rendus. Mécanique (2025), doi : 10.5802/crmeca.sp.1.

Les articles de cette revue sont mis à disposition sous la licence
Creative Commons Attribution 4.0 International (CC-BY 4.0)
https://creativecommons.org/licenses/by/4.0/deed.en

Comptes Rendus
de l'Académie des sciences

Mécanique

Recent advances in general relativity /
Avancées récentes dans le domaine de la relativité générale

Contents

Foreword to the special issue "Recent advances in general relativity: an issue in memory of Yvonne Choquet-Bruhat"
Jérémie Szeftel

On Yvonne
Piotr T. Chrusciel

A survey on the positive mass theorem for asymptotically flat initial data sets
Lan-Hsuan Huang

Gluing for the Einstein constraint equations
Justin Corvino

The stability of Minkowski space and its influence on the mathematical analysis of General Relativity
Jacques Smulevici

The black hole stability problem
Sergiu Klainerman

On the regularity problems of Einstein equations
Qian Wang

Weak cosmic censorship, trapped surfaces, and naked singularities for the Einstein vacuum equations
Yakov Shlapentokh-Rothman

The Strong Cosmic Censorship conjecture
Maxime Van de Moortel

Cosmology, the big bang and the BKL conjecture
Hans Ringström

Burnett's conjecture in general relativity
Arthur Touati

Some aspects of spectral and microlocal methods in Mathematical General Relativity
Dietrich Häfner

Cover illustration

Y. Choquet-Bruhat in her office at the Institut des Hautes Études Scientifiques in 2006 (photo by Jean-François Dars), in *Les Déchiffreurs : voyage en mathématiques* (ed. A. Lesne, J.-F. Dars and A. Papillault), Belin, 2008, p. 46.

1873-7234 (electronic)

Académie des sciences - Paris

Comptes Rendus
de l'Académie des sciences

Mécanique

An issue in memory of Yvonne Choquet-Bruhat - 2025