MIHALIS DAFERMOS

The formation of black holes in general relativity [after D. Christodoulou]

Astérisque, tome 352 (2013), Séminaire Bourbaki,
exp. n° 1051, p. 243-313

<http://www.numdam.org/item?id=AST_2013__352__243_0>

Séminaire BOURBAKI Mars 2012
64e année, 2011-2012, n° 1051, p. 243 à 313

THE FORMATION OF BLACK HOLES IN GENERAL RELATIVITY [after D. Christodoulou]

by **Mihalis DAFERMOS**

INTRODUCTION

Few notions of mathematical physics capture the imagination like that of *black hole.*

The concept was first encountered in explicit solutions of the Einstein vacuum equations:

$$\mathrm{Ric}(g) = 0, \tag{1}$$

specifically the celebrated *Schwarzschild* solution $(\mathcal{M}, g)_{\mathrm{Schw}}$. As was understood already by Lemaitre **[28]**, there is a region $\mathcal{B} \subset \mathcal{M}_{\mathrm{Schw}}$ of this spacetime with the property that observers in $\mathcal{B}$ cannot send signals to "far-away" observers. Following J. Wheeler, this region $\mathcal{B}$ is known as the *black hole* (or, in French translation, *le trou noir*).

More generally (and more precisely), one defines the black hole region of an asymptotically flat spacetime $(\mathcal{M}, g)$ as the collection of spacetime points $\mathcal{B} \subset \mathcal{M}$ *not* in the causal past of an ideal conformal boundary at infinity, so-called *future null infinity*, traditionally denoted $\mathcal{I}^+$; in symbols:

$$\mathcal{B} = \mathcal{M} \setminus J^-(\mathcal{I}^+). \tag{2}$$

In the explicit examples of Schwarzschild or Kerr, the existence of a black hole region $\mathcal{B}$ is accompanied by another salient feature: Every timelike or null geodesic $\gamma(s)$ entering the interior of $\mathcal{B}$ is future incomplete. In particular,

$$(\mathcal{M}, g) \text{ is } \textit{future-causally geodesically incomplete}. \tag{3}$$

In the case of Schwarzschild, the curvature grows without bound along *all* incomplete $\gamma(s)$ as the affine parameter s approaches its supremum value. In Kerr, the origin of incompleteness is in some sense even more bizarre; it represents not the breakdown of local regularity but of global causality.

Trapped surfaces and the theorems of Penrose

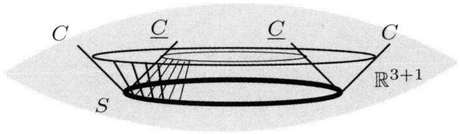
The physical implications of the above two properties (2), (3) are profound. Historically, however, they were very difficult to accept. A common point of view in the early years of the development of general relativity was to bet on an obvious way out of dealing with their consequences:

> *Could it be that the above "black hole"* (2) *and incompleteness properties* (3) *are pathologies, due to the high degree of symmetry of explicit solutions?*

One of the great successes of the global geometrical approach first pioneered by Penrose in the 1960s, was that it provided a definitive answer to the above question *in the negative.*

The key to this answer is provided by a fundamental notion introduced by Penrose [**34**], that of a *closed trapped surface.*

To motivate this notion, let us begin for sake of comparison with a standard 2-sphere S in Minkowski space $\mathbb{R}^{3+1}$ of radius 1.



If we consider the future of S, denoted $J^+(S)$, its boundary in $\mathbb{R}^{3+1}$ has two connected components, the two null cones C and $\underline{C}$ depicted. Considering the second fundamental form χ, $\underline{\chi}$ of S viewed as a hypersurface in each of the above null cones, respectively, we have that

$$\mathrm{tr}\underline{\chi} = -2 < 0, \qquad \mathrm{tr}\chi = 2 > 0.$$

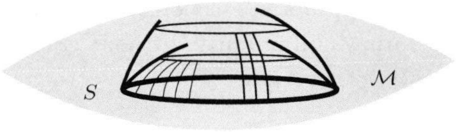
We call $\mathrm{tr}\underline{\chi}$, $\mathrm{tr}\chi$ the *future expansions* because they measure the change in the area element of the flow of S along the null generators of the respective cones.

Given now a general 4-dimensional time-oriented Lorentzian manifold $(\mathcal{M}, g)$, and a closed two surface S, we may again define the two second fundamental forms $\underline{\chi}$ and χ corresponding to viewing S as a hypersurface in each of the two connected components of the boundary of $J^+(S)$ intersected with a tubular neighborhood of S in $\mathcal{M}$. (Again, these are null hypersurfaces generated by the two sets of null geodesics orthogonal to S.)

We say that S is *trapped* if **both** its future expansions are negative:

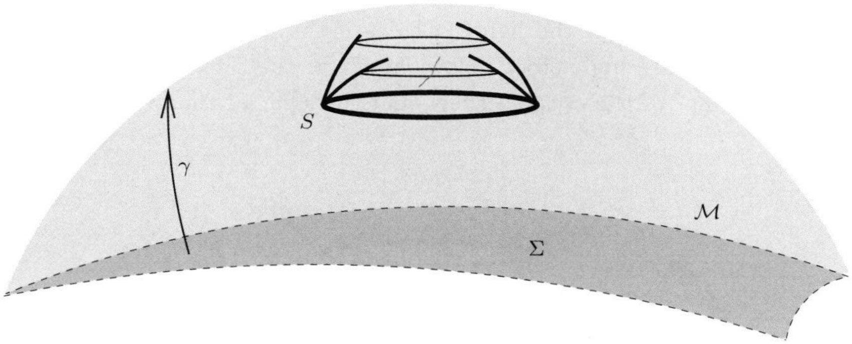
$$\mathrm{tr}\underline{\chi} < 0, \qquad \mathrm{tr}\chi < 0.$$

This is depicted here:



Penrose's celebrated *incompleteness theorem* then states:

THEOREM 0.1 (Penrose, 1965). — *Let* $(\mathcal{M}, g)$ *be globally hyperbolic with a non-compact Cauchy hypersurface, and let* $\mathcal{M}$ *satisfy* $\mathrm{Ric}(V, V) \geq 0$ *for all null* V. *It follows that if* $\mathcal{M}$ *contains a closed trapped surface, then it is future causally geodesically complete.*



Future causal geodesic incompleteness means that there exists an inextendible future-directed timelike or null geodesic γ whose maximum affine parameter is bounded above.

Note of course that when the vacuum equations (1) are assumed, the condition $\mathrm{Ric}(V, V) \geq 0$ is trivially satisfied.

The Einstein vacuum equations have a well-posed Cauchy problem (see Section 1.1.4). The statement of the theorem is such that it can be immediately applied to the maximal Cauchy development of asymptotically flat vacuum initial data (*if it is assumed that the spacetime contains a trapped surface* S), since the assumption of global hyperbolicity holds (by fiat!) for Cauchy developments (see Section 1.1.5). Let us note that by Cauchy stability, the existence of a closed trapped surface is now manifestly a stable property under perturbation of initial data. We obtain in particular the following:

COROLLARY 0.2. — *Let* $(\Sigma, \bar{g}, K)$ *be a sufficiently small perturbation of Schwarzschild data for the Einstein vacuum equations* (1). *Then the maximal Cauchy development* $(\mathcal{M}, g)$ *contains a closed trapped surface* S *and is geodesically incomplete.*

Concerning the black hole property, again one can state a very general result, assuming that one can define an appropriate notion of conformal boundary "at infinity", denoted $\mathcal{I}^+$, representing future null infinity. Without going into the details of such a definition (see Section 1.3), let us state:

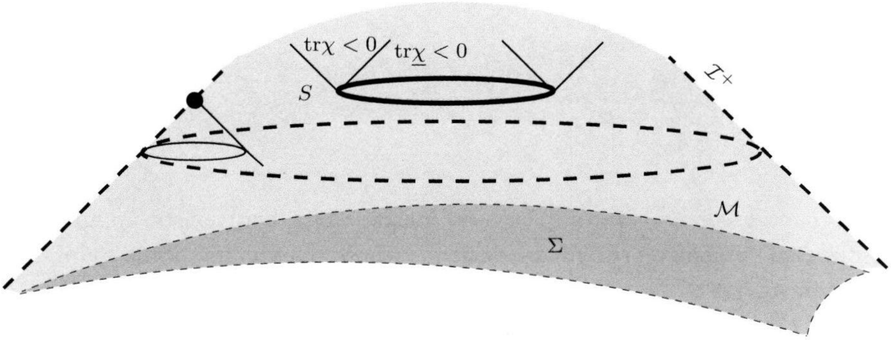
THEOREM 0.3 (See **[21, 42, 11]**). — *Under the assumptions of the previous theorem, if Σ is asymptotically flat and $\mathcal{I}^+$ is a suitable conformal boundary representing future null infinity, then*

$$S \cap J^-(\mathcal{I}^+) = \varnothing,$$

in particular

$$\mathcal{M} \setminus J^-(\mathcal{I}^+) \neq \varnothing. \tag{4}$$

Because this result makes reference to the causal structure, in depicting this, it is more appropriate to draw the light cones as if they are Minkowskian, so causal relations can be readily understood. Now, however, the area element is not to be inferred by the "size" of the cross-sections, and the signs of the expansions must be labelled:

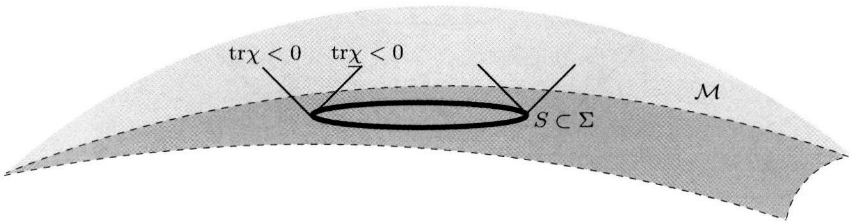


Our depictions in what follows will typically be of this form.

Similarly to Corollary 0.2, it follows that for sufficiently small perturbations of Schwarzschild data, the resulting Cauchy development $(\mathcal{M}, g)$ will still contain a (nontrivial) black hole region $\mathcal{B}$ (where the latter is defined simply in the sense of (2); see however Section 14!).

The main result: the dynamic formation of trapped surfaces

General as the above results may be, they do not shed light on whether black holes actually form *in nature*. The reason is that the assumption of the existence of a trapped surface is already a very strong assumption about the geometry of spacetime, one that—though stable—a priori may have nothing to do with properties of spacetimes that arise in physically interesting systems. Indeed, the only way previously known to ensure the assumption is to assume that there is a trapped surface S *already* in the initial data Σ:



or (as in Corollary 0.2) that one is close to Schwarzschild or Kerr (in which case, *any* initial data hypersurface—which would have to have 2 asymptotically flat ends!—contains points that are themselves contained in either a trapped surface or—worse!—a marginally anti-trapped[(1)] surface).

To link the above theorems with physical phenomena of gravity, one must answer the following question:

> *Do trapped surfaces form in evolution from the collapse of a regular (and arbitrarily dispersed) initial state, with trivial topology and geometry much like our own?*

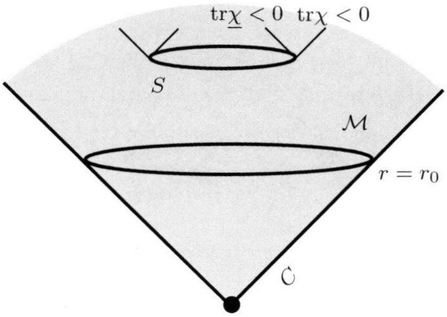
This lecture will report on a landmark result of Christodoulou—whose proof is published as a 580 page monograph **[9]**—which resolves the question posed above in the affirmative, for the Einstein vacuum equations (1)! Trapped surfaces form in evolution in vacuum collapse, in fact, they form from initial conditions which are infinitely dispersed.

THEOREM 0.4 (Christodoulou **[9]**, 2008). — *Trapped surfaces form in the Cauchy development of vacuum initial data which are arbitrarily (in fact infinitely!) dispersed.*

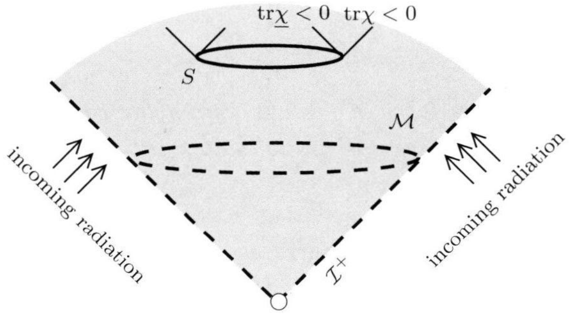
The formal statements of the above result are most naturally posed in terms of a characteristic initial value problem. One first formulates the problem on a finite initial cone where the data are in fact trivial up to a sphere of radius r_0. The quantity r_0 is

(1) An anti-trapped surface is one where the expansions χ and $\underline{\chi}$ are both positive. By marginally, we mean that positive here is to be taken in the French sense. Note that the names *future-trapped* and *past-trapped* are sometimes used for what we have called trapped and anti-trapped.

arbitrary and can be taken as a measure of the dispersion of the data.



See Theorem 2.1. The notion of "arbitrarily" dispersed can then be promoted to "infinitely" dispersed, by sending $r_0 \to \infty$. This can be formulated as the problem of prescribing scattering data "at past null infinity" $\mathscr{I}^-$, such that $\mathscr{I}^-$ is itself past complete.



Such data at past null infinity $\mathscr{I}^-$ has the interpretation of representing incoming radiation. This leads to the formulation in Theorem 13.1.

Outline of the exposition

This exposition will begin in Section 1 by reviewing those aspects of the Cauchy problem for the Einstein equations which are necessary for formulating these more precise versions of Theorem 0.4, recalling in particular the hyperbolicity of the equations, the well-posedness of the Cauchy and characteristic initial value problems, the proof of stability of Minkowski space, and its applications to the problem of gravitational radiation, in particular, rigorously defining a conformal boundary $\mathscr{I}^+$ at infinity. We will then proceed in Section 2 to give the statement of Theorem 2.1. This theorem in turn rests on a semi-global existence statement, Theorem 3.1, which is formulated in Section 3. Sections 4 to 10 then give details of the proof of Theorems 3.1 and 2.1. Section 11 gives a very brief exposition of a more recent approach to understanding the structure of the proof, due to Klainerman-Rodnianski **[27]**, leading to various simplifications. Applications to the incompleteness theorems are given in Section 12, and the formulation of the result with data at past null infinity $\mathscr{I}^-$ (Theorem 13.1) in Section 13. Finally, in Sections 14 and 15, the work is put in the context of open conjectures concerning black holes which are yet to be resolved!

(For the convenience of the reader, we note that various important formulae, equations, etc., are collected in a series of Appendices.)

The short pulse method and large data problems for non-linear wave equations

One should say at the onset that, general relativity aside, Theorem 0.4 is a "large-data", "large-time" result for a highly non-linear, supercritical system of hyperbolic PDEs, in more than 2 space-time dimensions, and such problems have hitherto for the most part been considered intractable.

From the PDE perspective, one of the main innovations of the work is the introduction of a new method for understanding large data results. This can be termed the "short pulse method".

In broad terms, the "short pulse method" seeks to introduce in a controlled way a certain large amplitude in the data, the *pulse*, compensated by a *short*ness in its characteristic length, the latter controlled by a parameter δ. This gives rise to a hierarchy of large and small quantities, which are coupled via the non-linear equations governing the theory. Given cooperative structural properties in the equations, one hopes that the evolution of data can be shown to respect this hierarchy, in essence because large amplitude parts are always appropriately coupled with sufficient smallness.

It is remarkable but in the end quite fitting that it is the Einstein equations (1), daunting in their complexity but so rich in structure, which have provided us with such a monumental realization of what promises to be a very fruitful general method to obtain large-data results, with many potential applications to nonlinear hyperbolic and other partial differential equations.

1. THE HYPERBOLIC AND RADIATIVE PROPERTIES OF THE EINSTEIN VACUUM EQUATIONS

As announced in the introduction, any discussion of the main result at a level more precise than Theorem 0.4 requires certain generalities about the Einstein equations to be reviewed. We thus introduce these very briefly in this section.

1.1. Hyperbolicity and the initial value problem

The most well-known aspect of the Einstein vacuum equations (1), which guided in particular Einstein to their discovery, is of course their "general covariance", or, in more modern language, their geometric content.

This miraculous geometric structure, however, in some sense obscures two other, equally fundamental aspects of (1):

The Einstein equations are hyperbolic and solutions radiate to infinity.

The relevance of hyperbolicity was already partially understood by Einstein in 1918, who was in fact the first to predict gravitational waves on the basis of his theory [**16**]. The conceptual issues involved, however, were confusing, and, as with several other issues in the theory, Einstein himself famously backtracked in the 1930's. The interested reader can consult [**22**].

1.1.1. *Harmonic coordinates.* — The most direct way to view the essential hyperbolicity of (1) is via the *harmonic gauge*, i.e. in coordinates x^μ which themselves satisfy the covariant wave equation

$$\Box_g x^\mu = 0, \tag{5}$$

where g is itself the spacetime metric. This is equivalent to requiring the following contraction of the Christoffel symbols to vanish:

$$\Gamma^\alpha_{\alpha\beta} = 0. \tag{6}$$

In such coordinates, the Einstein vacuum equations take the form

$$\Box_g g^{\mu\nu} = Q^{\mu\nu}(g, \nabla g) \tag{7}$$

where Q is quadratic in ∇g, and this is manifestly a system of quasilinear wave equations. This gauge was introduced by de Donder [**15**].

The equations (7) are known as the *reduced Einstein equations.* General uniqueness and existence statements for scalar equations of the form (7) were proven by Friedrichs-Lewy [**19**] and Schauder [**39**], respectively, using energy methods. Another approach to the analysis of equations of type (7) is through the construction of a parametrix (Hadamard, Petrovsky, Sobolev).

1.1.2. *The domain of dependence.* — The results on quasilinear wave equations of type (7) were given an immediate application to the Einstein equations by Stellmacher [**40**], who used the reduced equations (7) to prove a domain of dependence property for (1), capturing the important property of causality, in particular, uniqueness for smooth solutions to the initial value problem.

For the problem of existence, however, it is not immediate how to pass from (7) to (1)—it turns out that the necessary link is provided precisely by the so-called constraint equations, satisfied by initial data.

1.1.3. *Vacuum initial data and the constraint equations.* — The geometric formulation of initial data for (1) is provided by a triple

$$(\Sigma, \bar{g}, K),$$

where Σ is a 3-manifold to be a spacelike hypersurface in the evolving spacetime $(\mathcal{M}, g)$, and $\bar{g}$ and K are tensors to be the induced first and second fundamental forms, respectively, of Σ.

Whereas, however, initial data for the reduced Einstein equations (7) are free, initial data $(\Sigma, \bar{g}, K)$ for the actual Einstein equations are manifestly *not*! For, if $(\Sigma, \bar{g}, K)$ indeed imbedded in a Ricci flat manifold $(\mathcal{M}, g)$, then the contracted Gauss and Codazzi equations would imply

$$R_{\bar{g}} + (\mathrm{tr}K)^2 - |K|^2_{\bar{g}} = 0, \qquad \overline{\mathrm{div}}K - d\mathrm{tr}K = 0. \tag{8}$$

These are the so-called *Einstein constraint equations.*

1.1.4. *Local well-posedness for the Cauchy problem.* — The connection of (8) with the problem of passing from (7) to (1) was given in seminal work of Choquet-Bruhat [**17**].

Specifically, Choquet-Bruhat [**17**] proved that given a smooth solution of the constraint equations (8), then corresponding initial data could be set up for (7) such that (6) held to first order. Given local existence(2) for (7), the solution of (7) is then shown *a posteriori* to satisfy (6) because the quantity $\Gamma^\alpha_{\alpha\beta}$ itself satisfies a homogeneous wave equation (in view again of (7)). It follows that one has constructed a solution of the Einstein equations (1).

This resolves the question of local existence for (1).

1.1.5. *The maximal Cauchy development.* — In standard ODE theory for integral curves $\mathbf{x}(t)$ of a vector field, one can trivially "maximize" the local existence statement to infer the existence of a maximal solution, $\mathbf{x} : (T_-, T_+) \to \mathbb{R}^n$, where $-\infty \leq T_- < T_+ \leq \infty$. This is semantically useful as one can then talk about *the* solution.

In general relativity, there is a subtlety associated with this procedure, as the differential structure of spacetime is not given *a priori*, so there is not an obvious ordering on the domains on which solutions are defined.

(2) In [**17**], this is actually proven using a parametrix construction, but it can be obtained using the energy method of [**39**]. Most modern proofs essentially follow the method of [**39**], but parametrix constructions have found new applications. Cf. the references in Section 1.1.6.

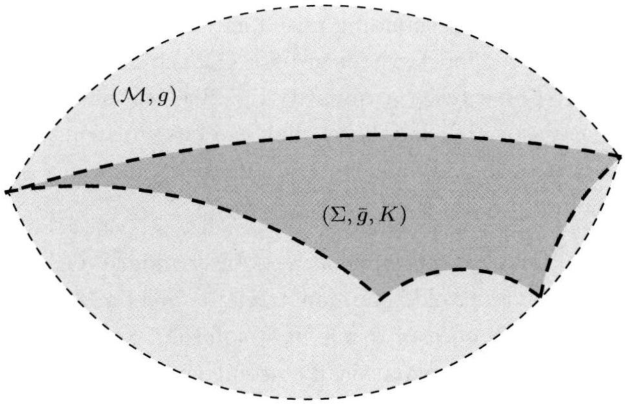
This was really only clarified in the late 1960's with the construction of the so-called maximal Cauchy development **[2]**(3). This notion depends on the concept of *global hyperbolicity*, introduced by Leray **[29]**, which ensures that the domain of dependence argument holds globally. Global hyperbolicity is the assumption that the spacetime admits a so-called *Cauchy hypersurface*, a hypersurface Σ with the property that every inextendible causal curve intersects Σ precisely once. The maximal Cauchy development is then uniquely characterized as the globally hyperbolic spacetime admitting the data, into which all other such spacetimes embed isometrically.

(Recall how the assumption of global hyperbolicity appears explicitly in the statement of the incompleteness theorem, Theorem 0.1, of the introduction.)

We summarize the well-posedness statement below:

THEOREM 1.1 (Choquet-Bruhat, Choquet-Bruhat-Geroch **[17, 2]**)

Let $(\Sigma, \bar{g}, K)$ be a smooth vacuum initial data set, i.e. a smooth solution of the Einstein vacuum constraint equations (8)*. Then there exists a unique, smooth maximal Cauchy development $(\mathcal{M}, g)$ satisfying the Einstein vacuum equations* (1)*, such that Σ is a Cauchy hypersurface in $\mathcal{M}$ with induced first and second fundamental form $\bar{g}$, K, respectively.*



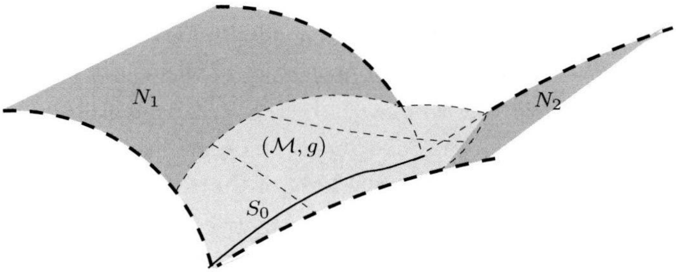
1.1.6. *Regularity.* — The above statement derives from a quantitative statement in spaces of finite differentiability. The question of proving existence and uniqueness in spaces of low regularity is one of much recent activity with many potential applications to the problem of singularity formation. We refer to **[25, 36]**.

(3) Interestingly, this construction, among other things, appeals to Zorn's lemma.

1.1.7. *The characteristic initial value problem.* — An alternative to the standard Cauchy problem is the characteristic initial value problem, where initial data is prescribed on either a null cone emanating from a vertex, or a bifurcate null hypersurface emanating from a spacelike 2-surface. As in the case of Cauchy data, a characteristic initial data set must satisfy constraints analogous to (8).

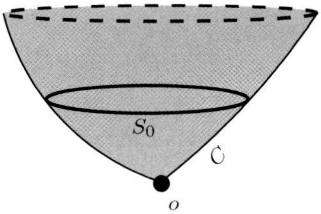
In the case of transversely intersecting initial null surfaces we have

THEOREM 1.2 (Rendall **[38]**). — *Consider an appropriate smooth vacuum characteristic initial data set defined on (what will be) null hypersurfaces N_1 and N_2 intersecting transversely on a spacelike surface $S_0 = N_1 \cap N_2$. There exists a non-empty maximal development $(\mathcal{M}, g)$ of the initial data bounded in the past by a neighborhood of S_0 in $N_1 \cup N_2$.*



The above theorem actually is proven by reducing the above problem in a clever way to the usual Cauchy problem.

In the case of initial data prescribed on a null cone C emanating from a vertex o,



then suitable conditions would need to be given on the vertex to ensure a solution. A particularly simple case is when the data are trivial up to a surface S_0. In this case, the question of existence reduces to Theorem 1.2, with N_1 the part of C to the future of S_0, and N_2 the trivial Minkowski ingoing cone $\underline{C}$ emanating from S_0. It is in fact in this form that we will appeal to the above in Section 2.

From one point of view, the characteristic initial value problem is more natural than the standard Cauchy problem, because a certain piece of the data can be specified freely, from which the constraints may then be solved simply by integration along the null generators of cones. This will become apparent in the present work, in fact we shall review in detail in Section 4 a very useful parametrization of this free data in terms of a non-standard geometric object ψ. (Cf. with the spacelike case, where the constraint equations (8) are a highly non-trivial underdetermined elliptic system.) On the other hand, although in the smooth category, Theorems 1.1 and 1.2 are analogous, the quantitative statement (expressed in Sobolev spaces of finite differentiability; cf. Section 1.1.6) behind well-posedness is different, as there is a fundamental loss in derivatives associated to the latter problem **[20]**.

A related issue is that the development $(\mathcal{M}, g)$ given by Theorem 1.2 does not contain a full neighborhood of the initial data hypersurface, only one of the 2-surface S_0. This problem has in fact been very recently overcome by Luk **[31]**.

1.2. The Bianchi equations

The above route to hyperbolicity, though adequate for the issue of well-posedness, does not capture precisely the *radiating* properties of the solution.

A much more geometric way to view both the *local hyperbolicity* properties, but also, the *global radiative* properties, is to view the wave-like features of the Einstein equations as fundamentally arising at the level of the *Bianchi equations.*

Recall that for a general pseudo-Riemannian metric, the Riemann curvature tensor satisfies the equations

$$\nabla_{[\epsilon,} R_{\alpha\beta]\gamma\delta} = 0. \tag{9}$$

If the manifold is four-dimensional and Ricci flat, i.e. if the Einstein vacuum equations (1) are satisfied, then one obtains also

$$\nabla^{\alpha} R_{\alpha\beta\gamma\delta} = 0. \tag{10}$$

The above set of equations (9)), (10 can be thought of as a generalization of the Maxwell equations and (in the Lorentzian signature) capture the basic hyperbolic and radiative properties of (1) in a geometric way.

1.3. The Newman-Penrose formalism, null infinity and peeling

Considerable insight in understanding these properties is gained by introducing a null frame and decomposing all relevant quantities with respect to that frame. This approach was originally pioneered by Newman-Penrose **[32]**. The structure equations of a general null frame coupled with the Bianchi identities written in terms of the frame is known as the *Newman-Penrose formalism.*

Using the above formalism, it was found that if one imposed the existence of a *smooth* (or at least C^2) conformal compactification of spacetime defining a boundary "future null infinity", denoted $\mathscr{I}^+$, then a certain hierarchy of fall-off immediately followed for various curvature components with respect to the frame. This hierarchy was in fact first discovered in **[35]** and is known as *peeling*.[(4)] This then allowed for the definition of rescaled Newman-Penrose curvature scalars, thought of now as functions *on* $\mathscr{I}^+$, which were in turn related to rescaled spin coefficients, the latter including the radiative amplitude, the quantity directly measured (cf. the definition (11) in Section 1.5). This gave a very powerful approach to defining and geometrically understanding the problem of gravitational radiation.

1.4. Stability of Minkowski space

The above approach to defining $\mathscr{I}^+$ and examining its properties, though extremely suggestive, rested on *a priori* assumptions on the global behavior of spacetime $\mathcal{M}$. In view of Theorem 1.1, however, the proper place for posing such assumptions is on *initial data* Σ.

At the time of **[32]**, it was in fact not clear that there were any examples of complete, one-ended initial data other than trivial data (i.e. data which develop to Minkowski space) leading to a spacetime for which one could define *any* reasonable notion of future null infinity $\mathscr{I}^+$.

The first (and essentially still only!) such problem which has since been globally understood is the case where initial data are a small perturbation of trivial data. This is the celebrated stability of Minkowski space **[12]**:

THEOREM 1.3 (Stability of Minkowski (Christodoulou-Klainerman, 1993))

For asymptotically flat initial data $(\Sigma, \bar{g}, K)$ satisfying a global smallness condition, the maximal vacuum Cauchy development $(\mathcal{M}, g)$ is geodesically complete, asymptotically approaches the flat metric in all directions, and admits complete asymptotic structures "future null infinity" $\mathscr{I}^+$ and "past null infinity" $\mathscr{I}^-$, such that moreover[(5)]

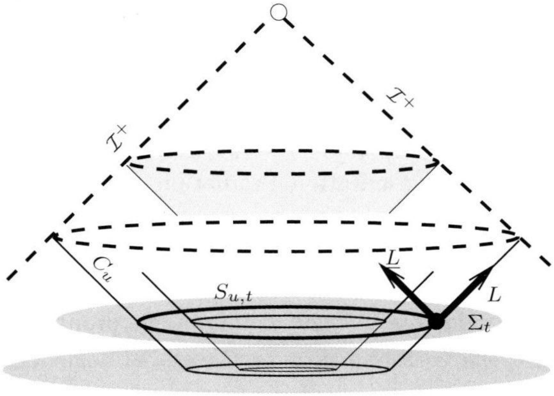
$$J^-(\mathscr{I}^+) = J^+(\mathscr{I}^-) = \mathcal{M}.$$

The monumental proof of the above theorem uses in part what can be thought of as a refinement of the Newman-Penrose formalism.

A fundamental insight of the work, however, is that the null frame should not be arbitrary but must be related to the choice of a suitable maximal time function t and

(4) In the language of the present work, this is in fact precisely the u_0-decay hierarchy in Section 5.4.3.
(5) In view of (2), this last statement is the assertion that the spacetime does not contain a black hole region (or its time reversed notion, namely white holes).

an optical function u, i.e. which together define two foliations Σ_t and C_u of spacetime, intersecting in 2 spheres $S_{u,t}$:



The asymptotic structure future null infinity $\mathscr{I}^+$ is constructed explicitly by "attaching" to each cone C_u a sphere at infinity with coordinate $\vartheta \in S^2$ defined via propagation along the null generators of C_u. The set $\mathscr{I}^+$ can then be viewed in the obvious way as a conformal boundary of $\mathcal{M}$. For the general data considered here, the regularity is only shown to be $C^{1,\alpha}$. See the comments below. Finally, the completeness statement is then that $\mathscr{I}^+$ is naturally parameterized by $(-\infty, \infty) \times S^2$.

The structure equations corresponding to an appropriate null frame—adapted to the above foliation—satisfy the so-called *optical equations*, which relate their connection coefficients and curvature. For example, the shear $\hat{\chi}$ of the cone C_u satisfies

$$\mathrm{d}\!\!/\mathrm{iv} - \frac{1}{2}\not{d}\,\mathrm{tr}\chi + \hat{\chi} \cdot \zeta^\sharp - \frac{1}{2}\zeta \cdot \mathrm{tr}\chi + \beta = 0.$$

These equations are much more powerful than the general structure equations of the Newman-Penrose formalism, because they can be used to capture a new global elliptic structure which had not been identified previously. This type of structure will play a role in the present work as well, as is discussed in Section 8.

In the proof of Theorem 1.3, the above structure had of course to be coupled with the hyperbolic, radiating aspects of the problem, captured at the level of energy estimates for the curvature tensor (which satisfies (9), (10)), as well as higher derivatives. These estimates are proven by an adaptation of the vector field method. This general idea will also be fundamental to the current work and will be discussed in Section 9.

Besides being a fundamental work in itself, Theorem 1.3 revealed several surprises about the nature of gravitational radiation, with bearing on the very physical tenability of the regularity assumptions on the conformal compactification described in Section 1.3. See in particular Sections 1.5.2 and 1.5.3 below.

Let us note that a variant of the foliation defined by a maximal time function t and an optical function u is to consider a double null foliation defined by two optical functions $\underline{u}$, u whose level sets are ingoing and outgoing cones. See **[10, 24]**. This will be in fact intimately connected to the framework of the present work.

Other extensions of stability of Minkowski space are contained in **[1]**. Finally, a harmonic gauge proof of a version of stability of Minkowski space is given in **[30]**.

1.5. Applications

1.5.1. *A general construction of* $\mathscr{I}^{\pm}$. — Essentially by domain of dependence arguments[6], one can show from Theorem 1.3 that given now an *arbitrary* asymptotically flat initial data set $(\Sigma, \bar{g}, K)$, not necessarily satisfying global smallness, in fact only assumed to be vacuum outside a compact set, one can still attach a *piece* of asymptotic boundary $\mathscr{I}^+$ to its Cauchy development, parameterized by $(-\infty, u_+) \times S^2$ such that, as before, $\{u\} \times S^2$ is "attached at infinity" to appropriate cones C_u.

Gravitational radiation is then described by limiting rescaled quantities, on $\mathscr{I}^+$.

Most fundamentally, one defines the radiative amplitude Ξ (per unit solid angle) to be the traceless symmetric two-tensor Ξ defined by

$$\Xi(u, \vartheta) = \lim_{C_u, r \to \infty} r \hat{\underline{\chi}}. \tag{11}$$

Similar considerations apply of course to $\mathscr{I}^-$. This, in particular, will allow one to formulate important physical conditions, such as "absence of incoming radiation". See in particular Theorem 1.4 in Section 1.5.3 below.

1.5.2. *Christodoulou memory.* — Gravitational wave experiments can be idealized as the study of the relative displacement of two test masses m_1, m_2 with respect to a third reference mass m_0, all located at large distance r and angular direction ϑ from a source. If the experiment is in the radiation zone of the source, then the whole configuration should be viewed instantaneously as sitting at a position $(u, \vartheta) \in \mathscr{I}^+$ for finite retarded time u, evolving in u. The u-rate of change of this relative displacement is determined precisely by the radiative amplitude $\Xi(u, \vartheta)$.

It turns out the Einstein equations at null infinity (See Chapter 17 of **[12]**) give the relation

$$\partial_u \Sigma \;= -\frac{1}{2}\Xi$$

where Σ is defined by

$$\Sigma(u, \vartheta) = \lim_{C_u, r \to \infty} r^2 \hat{\chi}. \tag{12}$$

[6] Alternatively, more directly by applying the double null foliation of **[10, 24]**.

Thus, for a suitable configuration of the test masses, the maximal change in displacement can be related to

$$\Sigma(u,\vartheta) - \Sigma^-(\vartheta) = -\frac{1}{2}\int_{-\infty}^{u} \Xi(u,\vartheta), \tag{13}$$

where

$$\Sigma^{\pm} = \lim_{u\to\pm\infty} \Sigma(u,\vartheta). \tag{14}$$

On the other hand, by examining the complete set of equations satisfied by the rescaled curvature components along null infinity $\mathscr{I}^+$, remarkably (see [4]), one can relate

$$\mathring{\mathrm{d}\!/\mathrm{iv}}(\Sigma^+ - \Sigma^-)$$

to

$$F(\vartheta) = \frac{1}{8}\int_{-\infty}^{\infty} |\Xi|^2(u,\vartheta)du, \tag{15}$$

more precisely to $F - \bar{F}$, where $\bar{F}$ denotes the mean over S^2. The expression (15) has the interpretation of total energy radiated per unit solid angle.

Using this expression, the *permanent* relative displacement, which is captured by

$$(\Sigma^+ - \Sigma^-)$$

can be related to (15).

For some physically interesting gravitationally radiating systems, it turns out that (15) can be of the same order of magnitude as the maximum in u of the right hand side of (13). See [4].

It follows that the total displacement of an appropriate configuration of test masses can be of the same order as the maximal displacement and may under certain circumstances be easier to measure. This effect is now known as the *Christodoulou memory effect.* See also the account by Kip Thorne [41].

1.5.3. *The end of peeling.* — A final achievement connected to the proof of the stability of Minkowski space was showing that the naive peeling properties suggested by the original analysis of [32] were in fact *not appropriate* for physically interesting solutions of the initial value problem.

This was proven in [8]. Since the statement is not widely known, we give a very short account here.

Recall by Section 1.5.1 above that the results of Theorem 1.3 concerning $\mathscr{I}^+$, $\mathscr{I}^-$ apply for general asymptotically flat data which are vacuum in a neighborhood of spatial infinity.

Recalling the radiative amplitude Ξ from (11), let us define, in analogy with (14),

$$\Xi^- = \lim_{u\to-\infty} u^2\Xi.$$

Note the weight in u. Finally, let us define the curvature coefficient β, with respect to a suitable null frame, as in Appendix 17.3. Peeling would have it that, for all cones C_u meeting $\mathscr{I}^+$, the asymptotic behavior of $\beta|_{C_u}$ is given by

$$\beta = O(r^{-4}). \tag{16}$$

To examine the appropriateness of (16), let us assume more generally that

$$\beta = B_*(u,\vartheta)r^{-4}\log r + O(r^{-4}), \tag{17}$$

for some $B_*(u,\vartheta)$ which possibly vanishes.

We have then the following remarkable result:

THEOREM 1.4 (Christodoulou [8], 1999). — *Let $(\mathcal{M}, g)$ be an asymptotically flat space time arising from a Cauchy hypersurface, vacuum in a neighborhood of spatial infinity.*

Assume that the Bondi mass is constant along past *null infinity $\mathscr{I}^-$ for advanced time $\underline{u}$ sufficiently large.*(7)

Then, if β on the cones C_u satisfies the asymptotics (17), *then B_* is given explicitly by the formula:*

$$B_* = -\frac{1}{4}(\mathring{\nabla\!\!\!/}\,\mathring{\mathrm{d}\!\!/\mathrm{iv}} + {}^*\mathring{\nabla\!\!\!/}\,\mathring{\mathrm{c}\!\!/\mathrm{url}})(\mathring{\mathrm{d}\!\!/\mathrm{iv}}\,\Xi^-). \tag{18}$$

(Note in particular that B_*, thought of as a function on $\mathscr{I}^+$, is independent of u.)

Thus, either

1. *The right hand side of* (18) *vanishes, or*
2. *$B_* \neq 0$ in* (17) *and peeling fails.*

The above theorem thus precisely identifies the origin of the logarithmic obstructions to peeling in terms of the asymptotic fall-off of the radiative amplitude Ξ as spacelike infinity is approached.

Why not try just to restrict to data which lead to a $\Xi^- = 0$? It turns out that Ξ^- has an interpretation in the post-Newtonian approximation in the simplest case of a spacetime arising from n particles falling from rest at infinity. In this context, the right hand side of (18) can be explicitly computed in terms of a moment of the configuration. In particular, B_* is indeed generically non-vanishing.

(7) This is precisely the assumption of "no-incoming radiation". Equivalently, one assumes that the analogue of Ξ defined on $\mathscr{I}^-$ vanishes identically for $\underline{u}$ sufficiently large.

Thus, generic physically admissible data will not exhibit peeling at future null infinity and the appropriate regularity for conformal compactifications is not better than $C^{1,\alpha}$.

2. MORE PRECISE STATEMENT OF THE RESULT

With this background on the Cauchy problem for the Einstein equations, we can now state a more precise version of Theorem 0.4. This will concern a statement for a finite characteristic initial value problem. The statement with data at past null infinity is deferred to Section 13.

2.1. The initial cone

The initial hypersurface will be a future null geodesic cone C_0 of a point o.

We will assume however that the initial data are trivial in a neighborhood of the vertex o.

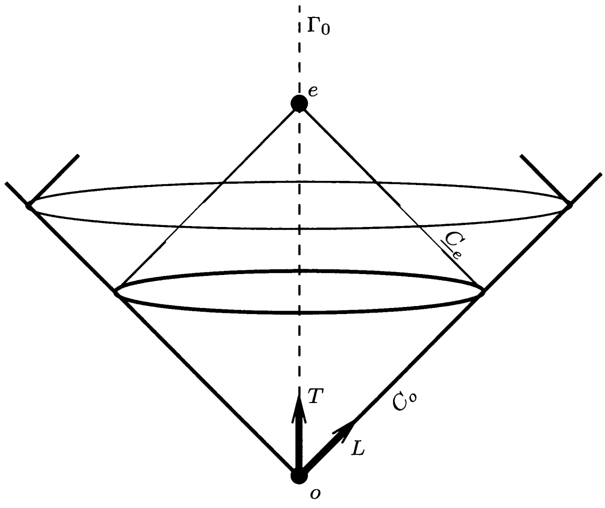
To describe the geometric interpretation of the data, as with the spacelike Cauchy problem, it is convenient to assume that we already have a spacetime into which the cone embeds.

Let T be a unit timelike future-directed vector at o and let Γ_0 be the geodesic generated by T. We can define the null vector L at o so as its projection to the span of T is T, and we can extend L along C_0 geodesically. Let s be the affine parameter defined by L.

Let $r_0 > 1$ be a constant.

Our assumption will thus be that the initial data are trivial for $s \leq r_0$, i.e. that they correspond to those of the corresponding truncated cone in Minkowski spacetime.

By the domain of dependence theorem, the solution will contain a Minkowski region which can be represented precisely as the past of a backwards light cone $\underline{C}_e$ emanating from a point e along Γ_0. Moreover, the length of the segment of Γ_0 connecting o and e is $2r_0$.



We will define an advanced time function $\underline{u}$ on C_o such that $\underline{u}|_{C_o} = s - r_0$. Later, this will be extended to the spacetime (to be constructed!) by the condition that its level sets are ingoing null hypersurfaces $\underline{C}_{\underline{u}}$. (See Section 3.1.)

Let us denote by $\not{g}$ the induced metric and χ the 2nd fundamental form of the sections of C_o corresponding to constant values of the affine parameter s. Let $\hat{\chi}$ denote the trace-free part of χ, and set

$$e = \frac{1}{2}|\hat{\chi}|^2_{\not{g}}. \tag{19}$$

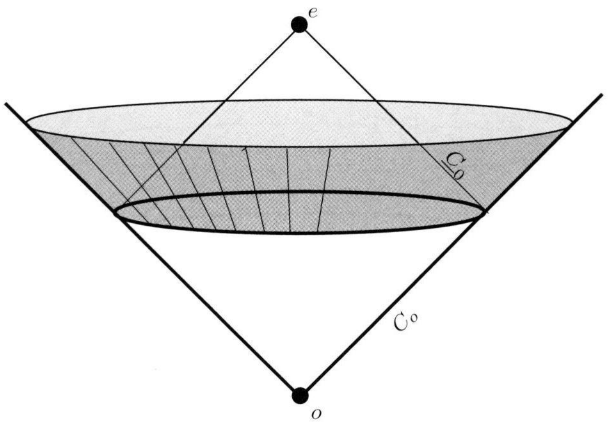
2.2. The trapped surface formation theorem

We may now state a version of the main theorem on the formation of trapped surfaces.

THEOREM 2.1 (Trapped surface formation). — *Let k, l be positive constants, $k > 1 > l$. For characteristic initial data as described above, suppose*

$$\frac{r_0^2}{8\pi}\int_0^\delta e\, d\underline{u} \geq \frac{k}{8\pi}, \tag{20}$$

where the integral is taken along the segment of each generator of C_o corresponding to the range $[0, \delta]$ of $\underline{u}$ for some $\delta > 0$.



Then, if δ is suitably small, the maximal development of the data contains a trapped sphere S of area

$$\text{Area}(S) \geq 4\pi l^2. \tag{21}$$

This function e actually only depends on the conformal class of $\not g$, which is in fact free data. We will discuss how to actually prescribe data in Sections 4 and 5. See in particular Section 5.3 for a discussion of the smallness condition on δ.

Remark 2.2. — The Einstein equations are clearly scale invariant under homotheties. Given $a > 0$, we may thus replace $r_0 \to ar_0$, $\delta \to a\delta$, $k \to ak$, $l \to al$ in the above statement. The formulation of the theorem has essentially set the scale to 1, and this corresponds to the order of the area radius of the trapped surface to be formed.

Remark 2.3. — As already announced in the introduction, we will extract in Section 13 a limiting statement from Theorem 2.1, where $r_0 \to \infty$ and C_0 is thus pushed to *past* null infinity $\mathscr{I}^-$. It should already be clear, however, that the function $r_0^2 e$ will correspond in the limit $r_0 \to \infty$ to the analogue of $|\Xi|^2$, as defined in (11), but now according to *past* null infinity, i.e. with χ in place of $\underline{\chi}$. Thus assumption (20) will correspond to a very natural condition on the flux of incoming radiation per unit solid angle. Cf. the role of this quantity in Section 1.5.2.

3. THE GAUGE AND THE SEMI-GLOBAL EXISTENCE THEOREM

The bulk of the work in obtaining Theorem 2.1 is proving a semi-global existence theorem. This will refer to a gauge defined by a double null foliation. The *existence* of the gauge will be *part of the theorem*, but as usual, it will be convenient to discuss its properties assuming it has already been shown to exist.

3.1. The double null foliation and canonical coordinates

3.1.1. *The optical functions u, $\underline{u}$.* — Recall the definition $\underline{u}|_{C_o} = s - r_0$ from Section 2.1. We shall extend $\underline{u}$ to the future of C_o so that its level sets are the future boundary of the part of the initial cone enclosed by the sets of constant $\underline{u}$ in C_o. We denote these level sets by $\underline{C}_{\underline{u}}$.

We introduce a function u conjugate to $\underline{u}$, the level sets of which are future null geodesic cones with vertices on Γ_0, and such that $u|_{\Gamma_0}$ measures arc length from o along Γ_0 minus r_0. We denote the level sets of u by C_u. We refer to u, $\underline{u}$ as our *optical functions.*

Finally, define the hypersurfaces H_t by $u + \underline{u} = t$.

In general, the foliation we are describing will only exist up to a $\underline{C}_\delta$, for some small δ and up to a hypersurface H_c as above, for $0 > c > u_0$. Let us denote such a region by M_c.

Let us make one final assumption on such an M_c: There are no cut or conjugate points along C_u, $\underline{C}_{\underline{u}}$ in $M_c \setminus \Gamma_0$.

Restricted to $M_c \setminus \Gamma_0$, we have then that C_u, $\underline{C}_{\underline{u}}$ are smooth null hypersurfaces, H_t defined above is a spacelike hypersurface, and $S_{\underline{u},u} = \underline{C}_{\underline{u}} \cap C_u$, are spacelike 2-surfaces diffeomorphic to S^2.

In what follows, all constructions will refer to some such M_c.

3.1.2. *The three null frames.* — We define first the null geodesic vector fields

$$\boxed{\underline{L}'^{\mu} = -2(g^{-1})^{\mu\nu}\partial_\nu \underline{u}, \qquad L'^{\mu} = -2(g^{-1})^{\mu\nu}\partial_\nu u.}$$

From these we define Ω by the relation:

$$-g(\underline{L}', L') = 2\Omega^{-2}.$$

Ω is the inverse density of the double null foliation.

We define two additional pairs, the *normalized pair*

$$\boxed{\hat{\underline{L}} = \Omega \underline{L}', \qquad \hat{L} = \Omega L'}$$

satisfying $-g(\hat{\underline{L}}, \hat{L}) = 2$ and finally the *equivariant pair*:

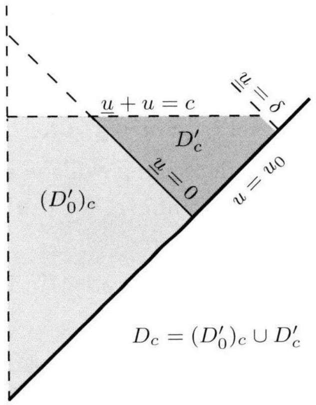
$$\boxed{\underline{L} = \Omega^2 \underline{L}', \qquad L = \Omega^2 L'}$$

satisfying $Lu = 0, \underline{L}\,\underline{u} = 0$, $L\underline{u} = \underline{L}u = 1$.

3.1.3. *Canonical coordinates.* — Let Φ_τ define the flow generated by L. Recall the definition $S_{\underline{u},u} = \underline{C}_{\underline{u}} \cap C_u$.

Note that $\Phi_\tau : S_{\underline{u},u} \to S_{\underline{u}+\tau,u}$ is a diffeomorphism. We can define similarly the flow $\underline{\Phi}_\tau$ generated by $\underline{L}$. Similarly $\underline{\Phi}_\tau : S_{\underline{u},u} \to S_{\underline{u},u+\tau}$ is a diffeomorphism.

Thus, given local coordinates $(\vartheta^1, \vartheta^2)$ in a patch $\mathcal{U}$ on the sphere S_{0,u_0}, we can extend these to $\underline{\Phi}_u(\Phi_{\underline{u}}(\mathcal{U})) \subset S_{\underline{u},u+u_0}$ by pullback. Thus, given two patches $\mathcal{U}_1$, $\mathcal{U}_2$ covering S_{0,u_0} with coordinates ϑ^A, $(\vartheta')^A$ respectively, the region $M_c \setminus \Gamma_0$ is covered by two coordinate patches with coordinates $(\underline{u}, u, \vartheta^1, \vartheta^2)$ and $(\underline{u}, u, (\vartheta')^1, (\vartheta')^2)$, and $(\underline{u}, \underline{u})$ range in a region D_c depicted below:



We may call these coordinate systems *canonical coordinates.*

Let us note that in such coordinates the metric takes the form

$$g = -2\Omega^2(du \otimes d\underline{u} + d\underline{u} \otimes du) + \not{g}_{AB}(d\vartheta^A - b^A d\underline{u}) \otimes (d\vartheta^B - b^B d\underline{u})$$

where b^A is governed by the torsion ζ (defined in Appendix 17.2.2) by the relation

$$\frac{\partial b^A}{\partial u} = 4\Omega^2 \zeta^{\sharp A}$$

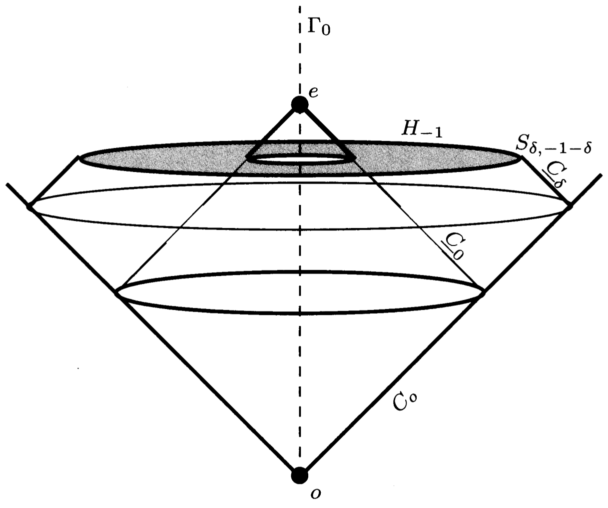
expressed in canonical coordinates.

For the problem of prescribing initial data, it will be convenient to choose a particularly nice coordinate system on the sphere, namely stereographic coordinates. We refer to Section 2.1 of [**9**] for the details.

3.2. The existence theorem

We may now state the semi-global existence theorem which is the heart of the work:

THEOREM 3.1 (Existence theorem). — *Consider data as described in Section 2.1. If δ is suitably small, then the maximal vacuum development of the data contains a region M_{-1} on which the gauge described in Section 3.1 can be constructed, bounded in the future by the spacelike hypersurface H_{-1} and the incoming null hypersurface $\underline{C}_\delta$, such that the cones $C_{\underline{u}}$ and C_u do not contain cut or conjugate points.*



Remark 3.2. — Under the assumptions of the trapped surface formation Theorem 2.1, the surface $S_{\delta,-1-\delta}$ will in fact be trapped, as will, by continuity, all surfaces $S_{\underline{u},u}$ with $\underline{u}$, u sufficiently close to δ, $-1-\delta$, respectively. As previously announced, we shall discuss the smallness assumption on δ in Section 5.3. Let us emphasize that indeed, the smallness assumption can be satisfied simultaneously in both Theorems 2.1 and 3.1.

4. FREE DATA

Before examining more closely the issue of smallness of δ and its significance, we must address the question: How does one actually parametrize the data?

We have already remarked that free data for the vacuum Einstein equations (1) in this context is precisely the conformal geometry of the initial cone. It will be essential, however, to have an explicit convenient parametrization of the space of free data. This will be provided by the geometric object $\psi\!\!\!/$ to be described immediately below.

4.1. The geometric object $\psi\!\!\!/$

We proceed with how to explicitly isolate the conformal geometry of the null cone with respect to our gauge.

4.1.1. *The conformal factor ϕ.* — On C_{u_0} let us write

$$g\!\!\!/|_{S_{\underline{u},u_0}} = (\phi|_{S_{\underline{u},u_0}})^2 \hat{g\!\!\!/}_{S_{\underline{u},u_0}}$$

where

$$d\mu_{\Phi^*_{\underline{u}}\hat{g\!\!\!/}|_{S_{\underline{u},u_0}}} = d\mu_{g\!\!\!/|_{S_{0,u_0}}}.$$

It follows that we may write

$$d\mu_{g\!\!\!/|_{S_{\underline{u},u_0}}} = (\phi|_{S_{\underline{u},u_0}})^2 d\mu_{\hat{g\!\!\!/}|_{S_{\underline{u},u_0}}}$$

for a scalar function ϕ.

In terms of canonical coordinates, we have

$$g\!\!\!/|_{S_{0,u_0}} = |u_0|^2 \mathring{g\!\!\!/}$$

where $\mathring{g\!\!\!/}$ represents the metric of the standard sphere. We may then write

$$\sqrt{\det g\!\!\!/(\underline{u}, u_0, \vartheta)} = (\phi(\underline{u}, \vartheta))^2 |u_0|^2 \sqrt{\det \mathring{g\!\!\!/}(\vartheta)}$$

whence

$$\hat{g\!\!\!/}_{AB}(\underline{u}, u_0, \vartheta) = |u_0|^2 \sqrt{\det \mathring{g\!\!\!/}(\vartheta)}\; m_{AB}(\underline{u}, \vartheta)$$

where $\det m = 1$.

4.1.2. *m and $\psi\!\!\!/$.* — The object m is in fact a 2-covariant symmetric positive definite *tensor density* of weight -1. We can see its transformation rule explicitly: Consider two charts ϑ and ϑ' covering S^2. Writing $\vartheta' = f(\vartheta)$ we may express

$$\frac{\partial f^A}{\partial \vartheta^B}(\vartheta) = T^A_B(\vartheta).$$

We see then that

$$m(\vartheta) = |\det T(\vartheta)|^{-1} \tilde{T}(\vartheta) m'(\vartheta') T(\vartheta).$$

If we restrict attention to stereographic charts on S^2 (as discussed in Section 3.1.3), then the matrix $O = \frac{T}{|\det T|^{1/2}}$ is orthogonal, symmetric of determinant -1.

We can write the matrix

$$m = \begin{pmatrix} Z + X & Y \\ Y & Z - X \end{pmatrix}$$

where $Z^2 - X^2 - Y^2 = 1$, i.e., this corresponds to the upper hyperboloid H_1^+.

If we consider now the exponential map:

$$\exp : \hat{S} \to H_1^+, \tag{22}$$

where $\hat{S}$ denotes the space of symmetric trace-free matrices, this defines an analytic diffeomorphism.

We can thus express $m = \exp \psi$ where $\psi \in \hat{S}$

$$\psi = \begin{pmatrix} a & b \\ b & -a \end{pmatrix},$$

and ψ transforms as

$$\psi(\vartheta) = \tilde{O}(\vartheta)\psi'(\vartheta')O(\vartheta). \tag{23}$$

We shall see in the following section that indeed ψ determines the whole set of initial data, once the Einstein equations are imposed.

To be completely concrete, one should consider a pair (ψ, ψ') each defining a map $[0, \delta] \times \overline{D}_{2\rho}$ for a $\rho > 1$, where the latter denotes the stereographic disc of radius strictly greater than 2, where ψ and ψ' in the overlapping region transform as above.

It will often be useful, however, to suppress this and consider ψ as a single geometric object[8]

$$\psi(\underline{u}, \boldsymbol{\vartheta}) : [0, \delta] \times S^2 \to \hat{S}$$

with $\boldsymbol{\vartheta} \in S^2$.

4.2. Determining the rest of the data

Given an arbitrary choice of ψ, one can then determine the remaining data by imposing the Einstein equations. See Appendix 17.4—17.6.

4.2.1. *Determining e.* — From (90), noting that $\Omega = 1$ on C_{u_0}, we have

$$\frac{\partial \mathrm{tr}\chi}{\partial \underline{u}} = -|\chi|^2_{\not g}.$$

We obtain

$$\mathrm{tr}\chi = \frac{2}{\phi}\frac{\partial \phi}{\partial \underline{u}}, \tag{24}$$

$$\hat{\chi}_{AB} = \frac{1}{2}\phi^2 \frac{\partial \hat{\not g}_{AB}}{\partial \underline{u}}. \tag{25}$$

We thus have

$$e = \frac{1}{2}|\hat{\chi}|^2_{\not g} = \frac{1}{8}(\hat{\not g}^{-1})^{AC}(\hat{\not g}^{-1})^{BD}\frac{\partial \hat{\not g}_{AB}}{\partial \underline{u}}\frac{\partial \hat{\not g}_{CD}}{\partial \underline{u}}.$$

[8] As such it is a funny object, namely, the logarithm of a tensor density of weight -1.

Note the function e is conformally invariant, i.e. it is independent of ϕ. Given, say stereographic coordinates, we may express it in terms of m and thus (inverting (22)) in terms of ψ.

4.2.2. *Determining* ϕ. — We have the linear equation

$$\frac{\partial^2\phi}{\partial^2\underline{u}} + e\phi = 0.$$

The initial conditions are

$$\phi|_{\underline{u}=0} = 1, \qquad \frac{\partial\phi}{\partial\underline{u}}|_{\underline{u}=0} = \frac{1}{2}\mathrm{tr}\chi|_{S_{0,u_0}} = \frac{1}{|u_0|}.$$

Thus, ϕ is determined in terms of e and thus of ψ.

Note that ϕ is a concave function. Let us call data *regular* if $\phi > 0$. This means precisely that C_{u_0} has no conjugate points. When we refer to "arbitrary data" in what follows, we shall always implicitly assume this condition.

4.2.3. *The rest.* — Continuing, one can obtain the initial values of all connection coefficients (see Section 17.2) and all curvature components (see Section 17.3) on C_{u_0}, in terms of ψ. For instance $\mathrm{tr}\chi$ is now defined by (24), $\hat{\chi}$ is defined by (25) (in view of the fact that $\not{g}_{AB}$ is defined by m which is defined by ψ), etc.

5. THE SHORT PULSE ANSATZ AND HIERARCHY

We have shown that free data is indeed completely parametrized by the geometric object ψ. In trying to understand the key to Theorems 2.1 and 3.1, it is useful to think of such ψ (restricted in $\underline{u}$ to $[0,\delta]$) as arising in a particular way. This will shed light both on the mechanism behind the result and on the quantities that determine the smallness of δ.

This way of defining ψ is what we shall call the *short pulse ansatz.*

5.1. The short pulse ansatz

Let ψ_0 be a "seed" map $\psi_0 : [0,1] \times S^2 \to \hat{S}$ which extends smoothly to $s < 0$.

Now given $u_0 < -1$, $0 < \delta < 1$, define for $\vartheta \in S^2$,

$$\psi(\underline{u}, \vartheta) = \frac{\delta^{1/2}}{|u_0|}\psi_0(\frac{\underline{u}}{\delta}, \vartheta). \tag{26}$$

Note that this assumption is indeed compatible with the transformation rule (23). This ψ defines now a mapping $[0,\delta] \times S^2 \to \hat{S}$.

5.2. Parametrization of initial data

One way of viewing the above ansatz is that for each fixed δ, it simply gives a specific way of parameterizing general data (restricting them in $\underline{u}$ to $[0,\delta]$).

That is to say, given initial data on C_{u_0} determined by ψ as in Section 4, and given δ, restricting ψ to $[0,\delta]\times S^2 \to \hat{S}$, we may *define* ψ_0 by (26).

The statements of the theorems will be more transparent if one thinks about fixing ψ_0 and defines ψ (and thus data on C_{u_0}) from ψ_0 after δ has been chosen; we shall freely move between both points of view.

5.3. The smallness assumption on δ

We are now ready to return to the issue of smallness of δ in the statement of Theorems 3.1 and 2.1. Small with respect to what? And are the assumptions simultaneously satisfiable?

Let us first take the point of view that we have given fixed data on C_{u_0} which determines ψ as in Section 4. Let us note that ψ must extend smoothly to 0.

Then, given δ, we may define ψ_0 as in Section 5.2 immediately above.

Let us now define M_k

$$M_k = \|\psi_0\|_{C^k([0,1]\times S^2)}. \tag{27}$$

The smallness condition on δ in both Theorems 2.1 and 3.1 can now be expressed as follows. There is a continuous, positive non-increasing function $F(M_8)$ such that

$$\delta < F(M_8). \tag{28}$$

In this approach where the data are first fixed, it follows that ψ_0 (and thus M_8) depends on δ. But in this approach, one sees that $M_8 \to 0$ as $\delta \to 0$. Thus, for fixed data, the smallness assumption (28) will indeed hold for sufficiently small δ.

Alternatively, one can begin by fixing ψ_0. Then M_8 is fixed, and clearly for δ sufficiently small, (28) holds. It follows that considering the rescaled data defined by ψ as in (26), then the statement of Theorem 3.1 holds.

The status of assumption (20) of Theorem 2.1 is somewhat different. For one sees that the quantity $\int_0^\delta e d\underline{u}$ is critical with respect to the rescaling defined by (26). Thus, (20) is essentially an assumption about the seed function ψ_0.

This assumption can be made more explicit by replacing (20) with the assumption

$$\frac{1}{8}\int_0^1 \left|\frac{\partial\psi_0}{\partial s}(s,\vartheta)\right|^2 \geq k, \qquad \frac{1}{8}\int_0^1 \left|\frac{\partial\psi_0'}{\partial s}(s,\vartheta)\right|^2 \geq k \tag{29}$$

(for all ϑ in two respective stereographic charts, as discussed in Section 4.1); this is directly computable from ψ_0 and it implies that (20) holds for all data rescaled by (26), with k replaced by $1+\frac{1}{2}(k-1)$. (See the statement of Theorem 7.1 of **[9]**.)

Let us note finally that the necessity of satisfying the transformation rule (23) is not an obstacle for satisfying both inequalities of (29).

Thus, in summary, starting with a seed function ψ_0 satisfying (29), then for δ satisfying (28), with $M_8 = M_8(\psi_0)$, then rescaled data determined by (26) will lead to developments satisfying the conclusions of both Theorems 3.1 and 2.1.

5.4. The initial data hierarchy

We fix now ψ_0 and will consider the corresponding initial data set on C_{u_0}, restricted to $\underline{u} \leq \delta$, which as we saw in Section 4.2 above, can be completely determined by defining ψ by (26). As before, we define M_k by (27) with respect to this fixed ψ_0.

One thing we have slipped under the rug is the question of whether, starting from ψ_0 and defining ψ by (26), one indeed obtains "regular" data, in the sense of Section 4.2.2. We note, however, that one easily sees that initial data are indeed regular if $\delta|u_0|^{-2}$ is sufficiently small depending on M_1.

5.4.1. *Invariant norms.* — Note first some basic control of the geometry: the eigenvalues of $|u_0|^{-2}\not{g}$ relative to $\mathring{\not{g}}$ are bounded above and below by fixed positive constants provided that δ is suitably small depending on M_1.

It follows that we can compare coordinate $|\cdot|$ norms, defined with respect to canonical coordinates in say two stereographic charts, with invariant norms: If ξ is a type T^q_p S-tensor field (see Appendix 17.1 for this notion!), then

$$C^{-1}_{p,q}|u_0|^{q-p}|\xi| \leq |\xi|_{\not{g}} \leq C_{p,q}|u_0|^{q-p}|\xi|.$$

Let us define invariant norms $\mathbf{C}^k_\delta$ by:

$$\|\xi\|_{\mathbf{C}^k_\delta} = \max_{m+n\leq k} \sup_{C_{u_0}} \left(\delta^n|u_0|^m|\nabla\!\!\!/^m D^n\xi|_{\not{g}}\right).$$

Now we say

$$\xi = \mathbf{M}_l(\delta^r|u_0|^s)$$

if for all k,

$$\|\xi\|_{\mathbf{C}^k_\delta} \leq \delta^r|u_0|^s F_k(M_{k+l})$$

for some F_k a nonnegative nondecreasing continuous function.

5.4.2. *The initial connection coefficient hierarchy.* — Just as one determined the data, one can now give bounds for the connection coefficients on the initial cone C_{u_0}, as these are defined in Appendix 17.2.

In the above notation, these bounds are as follows:

$$
\begin{aligned}
\hat{\chi} &= \mathbf{M}_1(\delta^{-1/2}|u_0|^{-1})\\
\mathrm{tr}\chi - \frac{2}{|u_0|} &= \mathbf{M}_1(|u_0|^{-2})\\
\zeta &= \mathbf{M}_2(\delta^{1/2}|u_0|^{-2})\\
\mathrm{tr}\underline{\chi} + \frac{2}{|u_0|} - 2\frac{\underline{u}}{|u_0|^2} &= \mathbf{M}_3(\delta|u_0|^{-3})\\
\underline{\hat{\chi}} &= \mathbf{M}_3(\delta^{1/2}|u_0|^{-2})\\
\underline{\omega} &= \mathbf{M}_3(\delta|u_0|^{-3}).
\end{aligned}
$$

5.4.3. *The curvature components.* — Recall that the curvature components in a null frame are denoted by the geometric objects α, β, ρ, σ, $\underline{\beta}$, $\underline{\alpha}$, defined in Appendix 17.3.

Similarly to the above, we obtain the following initial data hierarchy for curvature:

$$
\begin{aligned}
\alpha &= \mathbf{M}_2(\delta^{-3/2}|u_0|^{-1})\\
\beta &= \mathbf{M}_2(\delta^{-1/2}|u_0|^{-2})\\
\rho, \sigma &= \mathbf{M}_3(|u_0|^{-3})\\
\underline{\beta} &= \mathbf{M}_4(\delta|u_0|^{-4})\\
\underline{\alpha} &= \mathbf{M}_5(\delta^{3/2}|u_0|^{-5}).
\end{aligned}
\tag{30}
$$

5.4.4. *The nonlinearity of the hierarchy.* — We note that the above curvature hierarchy is nonlinear, in the sense that, had one used a linearized analysis, one would have obtained

$$\delta^{-3/2}, \qquad \delta^{-1/2}, \qquad \delta^{1/2}, \qquad \delta, \qquad \delta^{3/2}.$$

Let us note that to estimate correctly the latter two components $\underline{\beta}$, $\underline{\alpha}$ one must use the Bianchi identities (92) and (91) respectively. We see, for instance, that in the identity (91), the last two terms, though lower order from the point of view of differentiation, are dominant from the point of view of behavior in δ. This is a characteristic difficulty of the problem at hand.

Remark 5.1. — Let us note that the above non-linear δ-dependence of the curvature hierarchy is not unrelated to the non-linearity of the equations at infinity responsible for the memory effect discussed in Section 1.5.2. On the other hand, taking the limit as $u_0 \to \infty$ (see Section 13), we see that the u_0-dependence of the hierarchy reflects precisely the peeling hierarchy of **[32]**. This is still consistent, however, with the analysis of Section 1.5.3 concerning the validity of peeling. For as noted there, the analogue of B_*, defined with respect to past null infinity $\mathscr{I}^-$, is constant in $\underline{u}$ (as the roles of u and $\underline{u}$ are now reversed). Thus, it is the triviality of the data in the region $\underline{u} \leq 0$ that here imposes the analogue of $B_* = 0$, and thus peeling to indeed hold at past null infinity.

Remark 5.2. — Denoting by $\mathbf{M}_i$ the norms on the right hand side, it is also instructive to reflect upon the i-dependence of the hierarchy. One sees that although ψ is at the order of the metric and ζ, $\underline{\chi}$, $\underline{\omega}$ are at the order of first derivatives of the metric, the latter appear in Section 5.4.2 with $i = 3$. This is related to the loss of derivatives inherent to the characteristic initial value problem (cf. the remarks in Section 1.1.7). We shall not track this aspect of the hierarchy further in what follows.

5.5. Preservation of the hierarchy

The key to both the semi-global existence Theorem 3.1 and the trapped surface formation Theorem 2.1 is precisely the δ-hierarchy motivated by the above behavior of the norms in Sections 5.4.2 and 5.4.3 above.

This hierarchy will be encoded in the very energy estimates for curvature (and higher order so-called Weyl fields) that control the solution at top order. See Section 9.7. From there, it will filter down to all lower order estimates.

In particular, the connection coefficients will be bounded pointwise:

$$|\Omega| \leq O(1)$$

$$\left|\Omega \mathrm{tr}\underline{\chi} + \frac{2}{|u|}\right| \leq O(\delta|u|^{-2})$$

$$\left|\Omega \mathrm{tr}\chi - \frac{2}{|u|}\right| \leq O(|u|^{-2})$$

$$|\hat{\chi}| \leq O(\delta^{-1/2}|u|^{-1})$$

$$|\hat{\underline{\chi}}| \leq O(\delta^{1/2}|u|^{-2})$$

$$|\eta| \leq O(\delta^{-1/2}|u|^{-2})$$

$$|\underline{\omega}| \leq O(\delta|u|^{-3})$$

and curvature[9] will similarly be bounded pointwise:

$$|\alpha| \leq O(\delta^{-3/2}|u|^{-1}) \tag{31}$$

$$|\beta| \leq O(\delta^{1/2}|u|^{-2}) \tag{32}$$

[9] In examining the u dependence of (36), the reader may notice that a weaker bound is propagated than that suggested by the data (cf. (30)). This is because, although peeling can be shown to hold for smooth data of the type considered, the estimates at the level given do not propagate peeling. This is related to the actual failure of peeling for general initial data as discussed in Section 1.5.3.

$$|\rho| \leq O(|u|^3) \tag{33}$$

$$|\sigma| \leq O(|u|^3) \tag{34}$$

$$|\underline{\beta}| \leq O(\delta |u|^{-4}) \tag{35}$$

$$|\underline{\alpha}| \leq O(\delta^{3/2} |u|^{-9/2}). \tag{36}$$

Let us defer any further discussion of how these bounds are actually attained, and first turn briefly in the next section to showing that, *given* the semi-global existence Theorem 3.1 with respect to our gauge, *and* the propagation of the hierarchy in the form of the above bounds, we indeed obtain the result on formation of trapped surfaces, Theorem 2.1. In fact, adding

$$|\nabla\!\!\!/\, \eta| \leq O(|u|^{-3})$$

we will have written explicitly above precisely all those inequalities that we shall need!

6. PROOF OF THEOREM 2.1

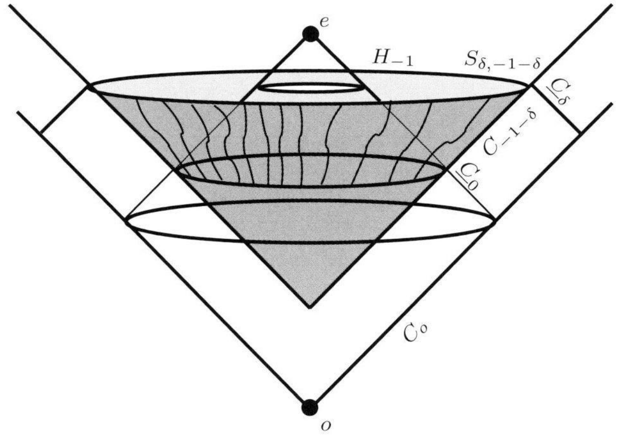
The proof of Theorem 2.1, *given* the semi-global existence Theorem 3.1 *and* the connection-coefficient estimates collected in Section 5.5, is in fact almost immediate, and we shall be able to essentially give the complete details in this short section.

6.1. Raychaudhuri on the cone $C_{-1-\delta}$

We will integrate the Raychaudhuri equation (90), written in the form

$$D\mathrm{tr}\chi' = -\frac{1}{2}(\mathrm{tr}\chi)^2 - |\hat{\chi}|^2$$

on the cone $C_{-1-\delta}$.



Defining

$$f = |u|^2 |\hat{\chi}|^2,$$

we have

$$D\mathrm{tr}\chi' \leq -(1+\delta)^{-2} f.$$

We note that

$$\mathrm{tr}\chi'|_{S_{0,-1-\delta}} = \frac{2}{1+\delta}$$

and thus, denoting by $\vartheta(\underline{u}, \vartheta_0)$ the ϑ-dependence of the null generators of $C_{-1-\delta}$ emanating from $(0, -1-\delta, \vartheta_0)$ in canonical coordinates as they cross $C_{\underline{u}}$, we obtain

$$\mathrm{tr}\chi'(\bar{u}, -1-\delta, \vartheta(\underline{u}; \vartheta_0)) \leq \frac{2}{1+\delta} - \frac{1}{(1+\delta)^2} \int_0^{\underline{u}} f(\underline{u}', -1-\delta, \vartheta(\underline{u}'; \vartheta_0)) d\underline{u}'.$$

It follows that if

$$\int_0^{\delta} f(\underline{u}, -1-\delta, \vartheta(\underline{u}; \vartheta_0)) d\underline{u} > 2(1+\delta) \tag{37}$$

for all $\vartheta_0 \in S^2$, then there is a $\underline{u}^* \in (0, \delta)$ such that for all $\underline{u} \in (\underline{u}^*, \delta)$ we have

$$\mathrm{tr}\chi'(\underline{u}, -1-\delta, \vartheta(\underline{u}; \vartheta_0)) < 0 \tag{38}$$

for all $\vartheta_0 \in S^2$, i.e. $S_{\underline{u},-1-\delta}$ is a trapped sphere.

6.2. Estimating the change in f from the short pulse hierarchy

We compute using the structure equation (88):

$$\underline{D} f = g \tag{39}$$

where

$$g = |u|^2 \left\{ -\left(\Omega \mathrm{tr}\underline{\chi} + \frac{2}{|u|} \right) |\mathrm{tr}\chi|^2 + 2(\hat{\chi}, \theta) \right\},$$

$$\theta = \Omega \left\{ \nabla\hat{\otimes}\eta + \eta\hat{\otimes}\eta - \frac{1}{2}\mathrm{tr}\chi\hat{\underline{\chi}} \right\} - \underline{\omega}\hat{\chi}.$$

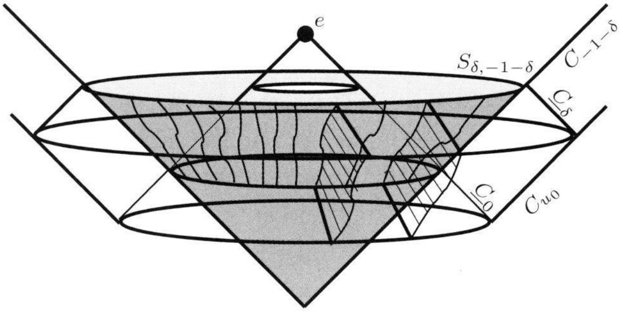
From the inequalities of Section 5.5, it follows that

$$|g| \leq O(\delta^{-1/2}|u|^{-2}). \tag{40}$$

Working in canonical coordinates, (39) has the form $\partial_u f = g$, and integrating we obtain

$$f(\underline{u}, -1-\delta, \vartheta) = f(\underline{u}, u_0, \vartheta) + \int_{u_0}^{-1-\delta} g(\underline{u}, u, \vartheta) du.$$

See



From (40), we have

$$-\int_{u_0}^{-1-\delta} |g(\underline{u}, u, \vartheta)| du \geq -O(\delta^{-1/2})$$

and so conclude that

$$f(\underline{u}, -1-\delta, \vartheta) \geq f(\underline{u}, u_0, \vartheta) - O(\delta^{-1/2}).$$

It follows that (37) is satisfied if

$$\int_0^\delta f(\underline{u}, u_0, \vartheta(\underline{u}; \vartheta_0)) > 2 + O(\delta^{1/2}). \tag{41}$$

6.3. The initial condition

Note that

$$f(\underline{u}, u_0, \vartheta) = 2|u_0|^2 e(\underline{u}, u_0, \vartheta) \tag{42}$$

for e defined in (19). In comparing (41) with (20) in view of (42), there is only one small subtlety remaining. The integration in (41) is not along the null generators of the cone C_{u_0}. This is of course a reflection of torsion. Nonetheless, using again the hierarchy (now concerning only the initial data as in Section 5.4) one can easily relate the two integrals modulo terms $O(\delta^{1/2})$, thus showing that (20) implies (37) and thus (38).

Let us note finally that the statement (21) concerning the area of the trapped sphere is again easily derived given the assumptions of the Theorem 2.1 and the estimates of the hierarchy given in Section 5.5.

7. THE PROOF OF THEOREM 3.1: A FIRST OVERVIEW

We now turn in the next three sections to the proof of the existence theorem, Theorem 3.1, and the intimately related property of the propagation of the hierarchy of Section 5.5.

As is typical for results concerning non-linear evolution equations, the proof is framed as a continuity argument, known in this context as a *bootstrap.*

For the benefit of the reader not familiar with such arguments, we will eventually want to outline in some detail how this is actually set up. It is hard, however, to motivate the ingredients of the set-up before one has introduced the main estimates of the proof, because it is the nature of the estimates themselves which define in particular the so-called bootstrap assumptions at the heart of the continuity argument.

We thus defer the outline of the actual continuity argument to Section 10. In Sections 8 and 9 below, we shall thus only consider the question of how to estimate a solution *assumed to exist* on a slab of spacetime of the form M_c, as in Section 3.1.

In broad terms, there are two parts to the problem of obtaining bounds, and this is already familiar from the proof of the stability of Minkowski space (see Section 1.4):

1. Use the structure equations to control the connection coefficients *given bounds on the curvature.*
2. Apply energy estimates to control curvature *given bounds on the connection coefficients.*

These two parts will reflect the breakdown between Sections 8 and 9, respectively.

8. CONTROLLING THE CONNECTION COEFFICIENTS FROM CURVATURE

In the present section we will review the structure which allows us to control the connection coefficients *given bounds on curvature.*

8.1. "Naive" propagation estimates

One may estimate "naively" the connection coefficients from curvature fluxes simply by integrating the propagation equations (78), (79), (80), (81), etc., along the null generators of the cones C_u, $\underline{C}_{\underline{u}}$.

This will indeed be used to derive L^∞ estimates from L^∞ estimates of the curvature (see Chapter 3 of [**9**]) and to derive $L^4(S)$[10] estimates for the first derivatives of the connection from $L^4(S)$ bounds for the first derivatives of the curvature (see Chapter 4).

For instance

$$|\hat{\chi}'| \leq C|u|^{-1}\delta^{-1/2}\mathcal{R}_0^\infty(\alpha) \tag{43}$$

[10] By this notation, we mean L^4 estimates with respect to the measure of the $S_{\underline{u},u}$ spheres.

where, in the notation of [9]:

$$\mathscr{R}_0^\infty(\alpha) = \sup_{M_c}(|u|\delta^{3/2}|\alpha|).$$

As is apparent, however, these estimates "lose" a derivative, as the connection coefficients are estimated at the same level as curvature, and not one degree better. The reason is clear: solutions of transport equations do not gain in differentiability against their right hand side. The presence therein of curvature thus necessitates the above loss.

These naive estimates are not however useless. In particular, although they lose a derivative, the estimates are sharp with respect to their δ-dependence; having such an estimate is important for the propagation of the hierarchy as in Section 5.5. Cf. Section 8.2.4.

8.2. A hidden elliptic structure

We will see in this section a hidden elliptic structure that will allow us to overcome the loss in regularity above. This has already appeared in the stability of Minkowski space, but as it is one of the most beautiful aspects of the structure of the Einstein equations, it is certainly worth repeating!

In the present work, this structure will be used specifically in order to obtain $L^4(S)$ estimates for second derivatives of χ, etc., from $L^4(S)$ estimates for first derivatives of curvature (see Chapter 6), and cone L^2 estimates for third derivatives of χ, etc., from cone L^2 estimates for second derivatives of curvature. The latter are top order estimates.

8.2.1. *Raychaudhuri and* $\mathrm{tr}\chi$. — Like almost everything in general relativity, the story begins with the Raychaudhuri equation (90)!

In its more general form (i.e. without imposing $\mathrm{Ric} = 0$):

$$D\mathrm{tr}\chi' = -\Omega^2|\chi'|^2 - \mathrm{Ric}(\hat{L}, \hat{L}), \tag{44}$$

this equation was first exploited by Penrose to prove his famous incompleteness theorem, Theorem 0.1. The significance of (44) in that context is that given the null curvature condition, then the right hand side of (44) is non-negative, and from this it follows that focal points form in finite time. From this in turn, geodesic incompleteness properties can be inferred by global topological methods.

Equation (44) plays a similar role in the proof of Theorem 0.3.

In the context of the analysis of the vacuum Einstein equations, the relation (44) takes on a new significance. The point is precisely that, in view of the vanishing of $\mathrm{Ric}(\hat{L}, \hat{L})$, no curvature term is present in (90).

In principle, this allows one to hope to estimate $\mathrm{tr}\chi$ without losing differentiability. Rewriting (90) in the form:

$$\mathrm{Dtr}\chi' = -\frac{1}{2}\Omega^2(\mathrm{tr}\underline{\chi}')^2 - \Omega^2|\hat{\underline{\chi}}'|^2, \tag{45}$$

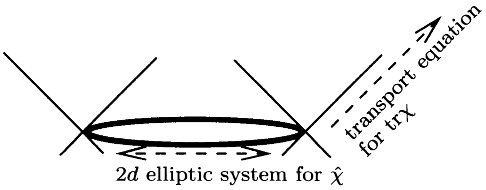
we see, however, that for this to work, one has to also estimate the trace-free part $\hat{\chi}$, the so-called *shear*, which appears on the right hand side of (45). (See Appendix 17.2.)

Here comes the second part of the miracle, which has no analogue in the classical use of (44) in the context of the proof of the incompleteness theorem.

Given control of $\mathrm{tr}\chi$, then the quantity $\hat{\chi}$, in view of the Codazzi equation (84), satisfies what can be viewed as an elliptic equation on the spheres $S_{\underline{u},u}$:

$$\slashed{\mathrm{div}}\,\hat{\chi}' = \frac{1}{2}\slashed{\mathrm{d}}\mathrm{tr}\chi' + \Omega^{-1}\left(-\beta + \frac{1}{2}\mathrm{tr}\chi\,\eta - \hat{\chi}^{\sharp}\cdot\eta\right). \tag{46}$$

The last two terms in brackets are clearly of lower order in differentiability. We see that one expects $\hat{\chi}'$ to have one more degree of angular regularity than the curvature form β and the form $\slashed{\mathrm{d}}\mathrm{tr}\chi'$. Thus, (46) coupled with (45)



allow one to obtain quantitative bounds for $\mathrm{tr}\chi$, $\hat{\chi}$, given bounds on curvature, gaining one level of differentiability.

Note that a similar argument can of course be applied for $\mathrm{tr}\underline{\chi}$, $\hat{\underline{\chi}}$.

8.2.2. *The mass aspect functions μ, $\underline{\mu}$.* — One would like to extend the above method to estimate the remaining connection coefficients $(\eta, \underline{\eta}, \omega, \underline{\omega})$. In examining the equations (80), (81), etc., from Appendix 17.4, however, it is not at all obvious how this is to be done, as there does not appear to be an immediate analogue of the good quantity $\mathrm{tr}\chi$ satisfying an analogue of the transport equation (45), with no curvature terms present on the right hand side.

One of the most beautiful discoveries associated with the original proof of the stability of Minkowski space **[12]**, is that there is indeed such a quantity!

The quantity is, however, at one level of differentiability higher than the connection coefficients! Specifically, one defines μ by

$$\mu = K + \frac{1}{4}\mathrm{tr}\chi\mathrm{tr}\underline{\chi} - \slashed{\mathrm{div}}\eta = -\rho + \frac{1}{2}(\hat{\chi}, \hat{\underline{\chi}}) - \slashed{\mathrm{div}}\eta. \tag{47}$$

We may also define a version associated with the $C_{\underline{u}}$

$$\underline{\mu} = K + \frac{1}{4}\mathrm{tr}\chi\mathrm{tr}\underline{\chi} - \not{\mathrm{d}}\mathrm{iv}\underline{\eta}.$$

We have that μ satisfies the transport equation

$$(48)\quad D\mu = -\Omega\mathrm{tr}\chi\mu - \frac{1}{2}\Omega\mathrm{tr}\chi\underline{\mu} + \Omega\left(-\frac{1}{4}\mathrm{tr}\underline{\chi}|\hat{\chi}|^2 + \frac{1}{2}\mathrm{tr}\chi|\underline{\eta}|^2\right) + \not{\mathrm{d}}\mathrm{iv}(\Omega(2\hat{\chi}^\sharp\cdot\eta - \mathrm{tr}\chi\,\underline{\eta})),$$

with a similar expression for $\underline{D\mu}$. This transport equation has the property that there are no curvature terms on the right hand side. To close, however, one still must somehow retrieve estimates for η, $\underline{\eta}$ and their derivatives on the spheres.

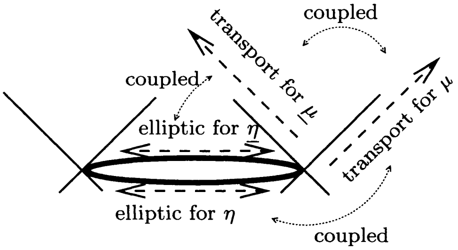
For this, recall the structure equation (82) from Appendix 17.4. Rewriting also (47) as an equation for $\not{\mathrm{d}}\mathrm{iv}\eta$, we see that for given μ, the quantity η can be viewed as satisfying the elliptic system

$$(49)\qquad \not{\mathrm{d}}\mathrm{iv}\eta = -\rho + \frac{1}{2}(\hat{\chi}, \underline{\hat{\chi}}) - \mu,$$

$$(50)\qquad \not{\mathrm{c}}\mathrm{url}\eta = \sigma - \frac{1}{2}\hat{\chi}\wedge\underline{\hat{\chi}}.$$

Similar considerations hold for $\underline{\eta}$ given $\underline{\mu}$.

We thus see that by simultaneously coupling both transport equations (48) and the equation for $\underline{D\mu}$, to the elliptic system (49)–(50) and the analogue for $\underline{\eta}$, we can in principle close the estimates.



The quantity μ has an important property. When integrated over the two-surface $S_{\underline{u},u}$, $\int_S \mu$ retrieves the so-called *Hawking mass*:

$$\sqrt{\frac{8\pi}{\mathrm{Area}(S_{\underline{u},u})}}\cdot m_{Hawk}(S_{\underline{u},u}) = 1 + \frac{1}{16\pi}\int_{S_{\underline{u},u}}\mathrm{tr}\chi\mathrm{tr}\underline{\chi} = \int_{S_{\underline{u},u}}\mu = \int_{S_{\underline{u},u}}\underline{\mu}.$$

With respect to the foliation considered in the proof of the stability of Minkowski space, the Hawking mass of $S_{t,u}$ approaches the so-called *Bondi mass* associated to the "cut" on $\mathscr{I}^+$ defined by the cone C_u. Moreover,

$$\lim_{t\to\infty}(\mathrm{Area}(S_{u,t})/4\pi)^{3/2}\underline{\mu}(u,t,\vartheta)$$

represents the energy per unit solid angle, the so-called news function $\underline{N}(u,\vartheta)$. See **[4]**, as well as Chapter 17 of **[12]**. Note that with Ξ defined by (11), we have

$$\frac{\partial \underline{N}}{\partial u} = -\frac{1}{2}|\Xi|^2.$$

Thus, replacing now future null infinity $\mathscr{I}^+$ with past null infinity $\mathscr{I}^-$ and exchanging the underlined and non-underlined quantities, we see that the change in N over a δ-interval along the generators of past null infinity measures precisely the limiting flux which appears in the statement of Theorem 13.1.

For this reason, μ and $\underline{\mu}$ are known as *mass aspect functions*, even though, strictly speaking, this interpretation is only valid at infinity.

8.2.3. *The quantity* $\underline{\psi}$. — For ω, $\underline{\omega}$, a suitable quantity is to be found one order further down in differentiability. We define

$$\underline{\psi} = \Delta\!\!\!/\,\underline{\omega} - \mathrm{d}\!\!\!/\mathrm{iv}(\Omega\underline{\beta})$$

to obtain the transport equation

$$D\underline{\psi} + \Omega\mathrm{tr}\chi\underline{\psi} = -2\Omega(\hat{\chi}, \nabla\!\!\!\!/^2\underline{\omega}) + \underline{m}$$

where $\underline{m}$ is of suitable order. For details, see Chapter 6.5 of **[9]**.

8.2.4. *Note on the δ-hierarchy.* — Without going into details, let us just remark that, in the case of η, $\underline{\eta}$, ω, the estimates obtained in the manner of the present section, though better from the point of view of differentiabilty, are worse with respect to δ that those obtained in Section 8.1 above.

The estimates are indeed sharp, however, in the case of χ, $\underline{\chi}$, and $\underline{\omega}$, and this is fundamental for the argument to close.

8.3. Sobolev inequality and uniformization on S

There is one missing link in order to apply elliptic theory on the $S_{\underline{u},u}$ spheres. For elliptic estimates to hold, one needs to first retrieve some basic underlying geometric information on the spheres.

First and foremost, to do the necessary elliptic theory, one requires the Sobolev inequality. This in turn is derived from the isoperimetric inequality

$$\int_{S_{\underline{u},u}} (f-\bar{f})^2 d\mu_{g\!\!\!/} \leq I(S_{\underline{u},u}) \left(\int_{S_{\underline{u},u}} |d\!\!\!/ f| d\mu_{g\!\!\!/} \right)^2$$

and $I(S_{\underline{u},u})$ is the isoperimetric constant. The latter can be estimated from the eigenvalues of the pullback $\Phi_u^* g_{S_{\underline{u},u}}$ with respect to $g|_{S_{0,u}}$. This in turn requires only a bound on

$$\delta|\Omega\mathrm{tr}\chi|, \qquad \delta|\Omega\hat{\chi}|.$$

The δ factor is quite fortuitous; comparing with the bounds of Section 5.5, there is in fact a $\delta^{1/2}$ to spare.

The other element which is required is the uniformization theorem. For in order to estimate in L^p, one needs to transfer Calderón-Zygmund estimates from the standard sphere to $S_{\underline{u},u}$.

In the context of stability of Minkowski space, a version of uniformization was proven **[12]** depending only on L^∞ bounds on curvature. In the present context, the L^∞ bounds on curvature are not sufficiently good for these purposes, in view of their bad δ-dependence.(11)

It turns out that uniformization theory can in fact be obtained using only $L^2(S)$ bounds on the Gauss curvature K. Moreover, from the equation

$$DK + \Omega \mathrm{tr}\chi K = \mathrm{d}\!\!/\mathrm{iv}\,\mathrm{d}\!\!/\mathrm{iv}(\Omega\chi) - \Delta\!\!\!/(\Omega \mathrm{tr}\chi),$$

we see immediately that an $L^2(S)$ estimate on K can be obtained from $L^2(S)$ flux type integrals of curvature. These flux integrals have the desired δ-dependence so as for the argument to close.

9. ENERGY ESTIMATES FOR CURVATURE

It is well-known that for hyperbolic equations in more than 2 spacetime dimensions, estimates for the top order quantities must be in L^2 on appropriate hypersurfaces, for all other norms would necessarily lose derivatives.

In this section we will describe how to prove such estimates for curvature and its first and second derivatives. It is the latter that will in fact be the *top order* quantities in this approach.

These estimates are of a geometric nature, and as we shall see will require precisely control on the connection coefficients in order to close. Thus, this and the previous section are in fact strongly coupled.

9.1. The energy method

Let us first briefly recall the energy method for the most classical example, a scalar field ψ.

We define the so-called *energy-momentum tensor* $\mathbb{T}[\psi]$ by

$$\mathbb{T}_{\mu\nu}[\psi] = \partial_\mu\psi\partial_\nu\psi - \frac{1}{2}g_{\mu\nu}\nabla^\lambda\psi\nabla_\lambda\psi.$$

(11) An analogous difficulty was encountered in Bieri's work **[1]**, where, in view of the fact that only H^1 bounds of curvature were being assumed on the initial Cauchy hypersurface, one could in principle only hope for $L^4(S)$ bounds on curvature. Thus, from pure regularity considerations, a version of uniformization was required assuming only $L^4(S)$-curvature bounds.

If ψ satisfies the covariant wave equation

$$\Box_g \psi = 0, \tag{51}$$

then

$$\nabla^\mu \mathbb{T}_{\mu\nu}[\psi] = 0.$$

More generally, for

$$\Box_g \psi = F,$$

we have

$$\nabla^\mu \mathbb{T}_{\mu\nu}[\psi] = F\partial_\nu \psi. \tag{52}$$

9.1.1. *Multiplier vector fields.* — Given a vector field V, we may now define a 1-form

$$P_\mu^V[\psi] = \mathbb{T}_{\mu\nu}[\psi]V^\nu$$

as well as a scalar current

$$\mathscr{J}^V[\psi] = {}^{(V)}\pi_{\mu\nu}\mathbb{T}^{\mu\nu}[\psi] + FV^\nu\partial_\nu\psi,$$

where ${}^{(V)}\pi^{\mu\nu}$ denotes the so-called deformation tensor of V, defined by

$${}^{(V)}\pi_{\mu\nu} = \mathscr{L}_V g_{\mu\nu}.$$

The relation (52) gives the divergence identity

$$\nabla^\mu P_\mu^V[\psi] = \mathscr{J}^V[\psi].$$

Applying the divergence theorem, we obtain an identity

$$\int_{\partial\mathscr{R}} P_\mu^V[\psi]n^\mu d\sigma_{\partial R} = \int_{\mathscr{R}} \mathscr{J}^V[\psi].$$

We note that the integrands of both boundary and bulk terms are of the same order of differentiability, quadratic in first derivatives.

If V is timelike, and $\partial\mathscr{R}$ consists of two homologous spacelike hypersurfaces Σ_0, Σ_1, then rewriting the above as:

$$\int_{\Sigma_1} P_\mu^V[\psi]n^\mu d\sigma_{\partial R} = \int_{\mathscr{R}} \mathscr{J}^V[\psi] + \int_{\Sigma_0} P_\mu^V[\psi]n^\mu d\sigma_{\partial R}, \tag{53}$$

the left hand side is now non-negative, by the general property

$$\mathbb{T}(V_1, V_2) \geq 0, \qquad V_1, V_2 \text{ causal, future-directed.} \tag{54}$$

It is in fact a coercive quantity in all derivatives of ψ, where the coercivity constants are determined by the geometry of Σ_1 and the choice of V.

If we consider $\mathscr{R}$ defined by a suitable foliation Σ_t, $t \in [0,1]$, then Gronwall's inequality allows one to use the above identity to estimate

$$|\psi|_{\mathring{H}^1(\Sigma_1)} \leq C|\psi|_{\mathring{H}^1(\Sigma_0)},$$

where C depends on the geometry of the foliation *and* the deformation tensor of the vector field V.

This estimate is particularly simple when V is Killing and $F = 0$, for then $\mathcal{J}^V = 0$. The relation (53) then represents Noether's theorem.

9.1.2. *Commutation vector fields.* — To obtain higher L^p bounds (including pointwise bounds), one must obtain energy estimates at a higher order of regularity. This is done by introducing so-called commutation vector fields.

Note that if X is a vector field and $\Box_g \psi = 0$, then

$$\Box_g (X\psi) = [\Box_g, X]\psi = F(\partial\psi, \partial^2\psi).$$

It is again the deformation tensor of X that enters in the coefficients of the expression F.

Given now also a multiplier vector field V, we may apply identity (53) to $X\psi$ to bound say

$$\int_{\Sigma_1} P^V_\mu [X\psi]\, n^\mu$$

and more generally

$$\int_{\Sigma_1} P^V_\mu [X_1 \cdots X_n \psi]\, n^\mu. \tag{55}$$

To relate the quantity (55) to a higher order Sobolev norm of ψ, one needs to examine the coercivity properties of the expressions in various directions. The most immediate way to obtain coercivity in given directions is to include commutation vector fields X_i which span those directions.[12] These higher order energy estimates in turn lead to pointwise or higher L^p estimates for lower order quantities via Sobolev imbedding type theorems.

9.1.3. *The vector field method.* — The combined use of multiplier and commutation vector fields is behind both the local well-posedness and long time decay properties for linear and non-linear wave equations. When used to prove decay results, both the multiplier and the commutation vector fields have well-chosen weights. See Klainerman's seminal **[23]** for how this approach captures the dispersive properties of (51) in a way sufficiently robust for applications to non-linear stability properties. This is the celebrated *vector-field method.*

[12] One can also, however, sometimes use the equation (51) itself together with elliptic estimates to control directions which are not in the span of the commutation vector fields.

9.2. The Bel Robinson tensor

In the present work, one wishes to apply a version of the vector field method at the level of the Bianchi equations for the Riemann curvature tensor (cf. the discussion in Section 1.2). From the perspective of its algebraic structure, the Riemann curvature tensor is a special case of an object known as a *Weyl field*, a subclass of the set of covariant 4-tensors. The algebra and calculus of such fields are reviewed in Appendix 17.7.

For a Weyl field W, we define the totally symmetric, trace free quadratic expression in W:

$$\mathbb{Q}_{\alpha\beta\gamma\delta}[W] = (W_{\alpha\mu\beta\nu}W_{\gamma\ \ \delta}^{\ \mu\ \ \nu} + {}^*W_{\alpha\mu\beta\nu}{}^*W_{\gamma\ \ \delta}^{\ \mu\ \ \nu})/2.$$

This is the celebrated *Bel Robinson tensor.*

If W satisfies the Bianchi equations (9), (10), then $\mathbb{Q}$ is divergence free

$$\nabla^\alpha \mathbb{Q}_{\alpha\beta\gamma\delta}[W] = 0.$$

More generally,

$$\mathrm{div}\mathbb{Q}\,[W] = (W,\mathcal{J}),$$

where $\mathcal{J}_{\beta\gamma\delta}$ is a so-called Weyl current defined by (93).

In analogy with (54), $\mathbb{Q}$ has the remarkable property

$$\mathbb{Q}(W)(V_1,V_2,V_3,V_4) \geq 0, \qquad V_i \text{ causal, future-directed.} \tag{56}$$

Thus $\mathbb{Q}$ can indeed be considered a close analogue to the $\mathbb{T}$ of Section 9.1.

9.3. Multiplier and commutation fields

The approach of using an adaptation of the classical vector field method with the Bel Robinson tensor $\mathbb{Q}$ in the role of the energy momentum tensor $\mathbb{T}$ was first taken in the proof of stability of Minkowski space **[12]**. This will be the approach used here to obtain estimates. In analogy with the classical vector field method, one must first discuss the choice of multipliers and commutators.

9.3.1. *Multiplier fields.* — Let us first remark that, given a Weyl form W, to obtain a 1-form from $\mathbb{Q}[W]$, in general we must contract with 3 vector field multipliers!

$$P^\alpha[W]^{(V_1,V_2,V_3)} = -\mathbb{Q}^\alpha_{\ \beta\gamma\delta}[W]V_1^\beta V_2^\gamma V_3^\delta.$$

How are these chosen in the present work?

As multiplier fields, we shall always choose from

$$L, \qquad K$$

where $K = u^2\underline{L}$.

Note that these are both causal, and thus, in view of (56), they generate non-negative definite boundary terms if the divergence theorem is applied in a region bounded by spacelike or null hypersurfaces with the appropriate orientations. The weight u^2 in K is essential to track the correct behavior with respect to u in the hierarchy.

9.3.2. *Commutation fields.* — As commutation fields we will take

$$L, \qquad S, \qquad O_i \qquad i = 1, 2, 3$$

where $S = \underline{u}L + u\underline{L}$, and O_i denote the so-called *rotation fields.*

The latter are defined most directly in canonical coordinates simply as the standard rotations of the sphere S^2. (See Chapter 8.3 of **[9]**.) These will satisfy coercivity properties to be discussed below in Section 9.5.

9.4. Deformation tensors and connection coefficients

As we shall see, just as in the classical vector field method, it is the deformation tensors of the multiplier and commutation vector fields which will appear in our energy identities. Thus, to estimate curvature, we will need in particular to estimate these deformation tensors.

Since the vector fields are related to the null foliation, it should not be surprising that our deformation tensors can in turn be estimated by connection coefficients. Indeed, the deformation tensors of our multiplier and commutator vector fields are estimated in Chapters 8 and 9 of **[9]** from precisely the type of quantities estimated in Section 8.

It is in this way that energy estimates are coupled to the estimates of Section 8. In particular, it is through this that the δ-hierarchy for connection coefficients (derived with the help of the curvature hierarchy) will re-couple with the estimates required to *prove* the propagation of the curvature hierarchy. We will discuss this in Section 9.10 in the context of the most difficult terms, the so-called *borderline terms.*

9.5. Aside: Sobolev inequalities and coercivity

In analogy with our discussion of the classical vector field method, we know that for energy identities to give control on geometrically natural L^2 quantities we need coercivity properties, e.g. properties of the form

$$\sum_i |\not{\mathcal{L}}_{O_i}\xi|^2 + |\xi|^2 \geq C^{-1}(|u|^2|\not\nabla\xi|^2 + |\xi|^2).$$

To moreover then get control on higher L^p, we need control on the relevant Sobolev constants, both on the spheres $S_{\underline{u},u}$ and the cones $C_{\underline{u}}$, C_u.

The issue of the Sobolev inequality on the spheres has in fact already been discussed in Section 8.3 and is addressed in Chapter 5 of **[9]**. The remaining issues are addressed in Chapters 10 and 11 of **[9]**.

Recalling the discussion in Section 8.3, we emphasize that control on coercivity, Sobolev, etc., require at the very least some basic assumptions about the geometry of the spheres and cones.

In the logic of the proof as described in Section 10, the basic bounds necessary for these must be included as bootstrap assumptions, and are finally recovered from the estimates proven. We have given already some flavour of the origin of this improvement in Section 8.3 above. We shall not comment further on this in the present section.

9.6. Table of Weyl fields used, and the notion of index

Before proceeding, we collect here the complete list of Weyl fields to be used:

Order	Weyl field W
0th order	R
1st order	$\tilde{\mathcal{L}}_L R$, $\tilde{\mathcal{L}}_{O_i} R$, $\tilde{\mathcal{L}}_S R$
2nd order	$\tilde{\mathcal{L}}_L\tilde{\mathcal{L}}_L R$, $\tilde{\mathcal{L}}_{O_i}\tilde{\mathcal{L}}_L R$, $\tilde{\mathcal{L}}_{O_i}\tilde{\mathcal{L}}_{O_j} R$, $\tilde{\mathcal{L}}_{O_i}\tilde{\mathcal{L}}_S R$, $\tilde{\mathcal{L}}_S\tilde{\mathcal{L}}_S R$

We will define the *index* ℓ of a Weyl field by the rule:

$$\ell(W) = \sharp \text{ of } \tilde{\mathcal{L}}_L \text{ operators in the above representation.}$$

The significance of the index will become clear in Section 9.7 below.

9.7. The (n)-stratification and the exponents of the short pulse hierarchy

We now turn to what essentially is the heart of the whole method.

We wish to capture at energy level the curvature hierarchy of Section 5.5. It turns out that the different δ-behavior of the various curvature components α, β, is distinguished precisely by the freedom of choosing different combinations of *multiplier* vector fields.

This motivates the following: Given a Weyl field W, we define $\overset{(n)}{P}[W]$ for $n = 0, 1, 2, 3$ by

$$\overset{(0)}{P}[W] = P[W](L, L, L),$$
$$\overset{(1)}{P}[W] = P[W](K, L, L),$$
$$\overset{(2)}{P}[W] = P[W](K, K, L),$$
$$\overset{(3)}{P}[W] = P[W](K, K, K).$$

The (n) above measures the number of K's used as multipliers.

Let us consider the case $W = R$. The point is now that for each n, different combinations of curvature components appear in the flux terms associated to $\overset{(n)}{P}[R]$. Thus, n provides a stratification of currents that will allow one to implement the curvature hierarchy of Section 5.5. To be consistent with this hierarchy, it turns out that one must "assign" δ exponents to each n by the rule:

$$q_0 = 1, \qquad q_1 = 0, \qquad q_2 = -1/2, \qquad q_3 = -3/2. \tag{57}$$

We shall refer to the above q_n as the *exponents* q_n of the short pulse hierarchy.

This assignment essentially captures the fact that, *when applied to curvature* $W = R$, we expect the boundary terms associated to $\overset{(n)}{P}[R]$ in the energy identity to be like δ^{-2q_n}.

Of course, we must apply currents not only with $W = R$ but with higher order Weyl fields from the last two rows of our table in Section 9.6 above. But now, because $\tilde{\mathscr{L}}_L$ differentiates in the $\underline{u}$ direction, we expect quadratic quantities with index ℓ to be $\delta^{-2\ell}$ worse than they would be otherwise.

This motivates the definition:

$$\overset{(n)}{P} = \sum_W \delta^{2\ell} \overset{(n)}{P}(W). \tag{58}$$

(In the case $n = 3$ we sum over all Weyl fields from our table in Section 9.6. In the cases $n = 1, 2$, we omit from the sum those containing an $\tilde{\mathscr{L}}_S$ in their representation.)

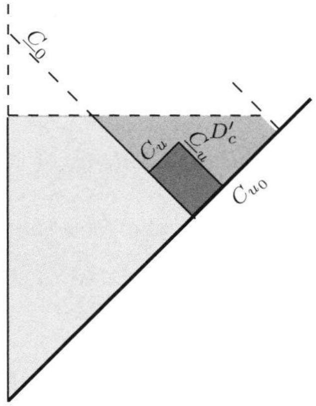
With the inclusion of the index factor, all terms in the sum (58) have the same expected δ-behavior. In view of the above comments, we expect the boundary terms of $\overset{(n)}{P}$ in the energy identity to be like δ^{-2q_n}.

9.8. The divergence identity

In the previous section, we have successfully translated the δ-hierachy of curvature components in Section 5.5 into a δ-hierarchy for energy currents $\overset{(n)}{P}$, stratified by (n), and given by (57).

We are now ready to apply the divergence theorem for each $\overset{(n)}{P}$, $n = 0, \dots, 3$.

In the notation of (3.1), we will consider a domain as below



and integrate the divergence identity

$$\operatorname{div} \overset{(n)}{P} = \overset{(n)}{\tau} \tag{59}$$

over the darker shaded region (where this is to be envisioned rotated around Γ_0).

Denoting by E the boundary term on C_u, F the boundary term on $\underline{C}_{\underline{u}}$, and D the boundary term on C_{u_0} ("data"), we may write the resulting identity as:

$$E(u) + F(\underline{u}) = \int \tau d\mu_g + D(\underline{u}). \tag{60}$$

In view of our hierarchy (57), let us define

$$\overset{(n)}{\mathscr{E}} = \sup_u \left(\delta^{2q_n} \overset{(n)}{E}(u) \right),$$

$$\overset{(n)}{\mathscr{F}} = \sup_{\underline{u}} \left(\delta^{2q_n} \overset{(n)}{F}(\underline{u}) \right),$$

the data terms:

$$\overset{(n)}{D} = \delta^{2q_n} \overset{(n)}{D}(u_0),$$

and finally

$$\mathscr{P} = \max\{ \overset{(0)}{\mathscr{E}}, \overset{(1)}{\mathscr{E}}, \overset{(2)}{\mathscr{E}}, \overset{(3)}{\mathscr{E}}, \overset{(3)}{\mathscr{F}} \}. \tag{61}$$

Our goal is thus to bound $\mathscr{P}$ using (60).

9.9. The excess index

In view of (59) and the above definitions, we must estimate

$$\delta^{2q_n} \int |\overset{(n)}{\tau}| d\mu_g \leq O(\delta^{2e}) \tag{62}$$

where e is the so-called *excess index.*

If it were always the case that $e > 0$, then the estimates would immediately close for sufficiently small δ. For irrespectively of the non-linear dependence of the right hand side of (62) in $\mathscr{P}$, we would have a smallness factor provided by δ^{2e}.

On the other hand, if $e < 0$, then it would be hopeless to prove estimates, and the hierarchy would essentially be inconsistent.

9.10. The borderline terms

The good news is that it turns out that

$$e \geq 0.$$

The bad news, however, is that there are terms which appear on the right hand side of (59) for which one must take $e = 0$ in (62)

These are the so-called *borderline terms.*

To see these, let us examine the structure more carefully:

For each choice of multiplier set and Weyl field W we have the identity

$$\begin{aligned} \mathrm{div} P[W] &= -(\mathrm{div} Q)[W](V_1, V_2, V_3) - Q^{\alpha\beta}_{\gamma\delta}[W]\, {}^{(V_2)}\pi_{\alpha\beta} V_1^\gamma V_3^\delta + \cdots \\ &= \tau_c + \tau_m. \end{aligned}$$

For the first term above, recall

$$\mathrm{div} Q = (W, \mathscr{J}).$$

For the fundamental Weyl field, namely curvature, $\mathscr{J} = 0$. We see thus that the first term only arises because of commutation fields. The error term τ_c is thus introduced by the commutation fields, while the error term τ_m is introduced by the multiplier fields.

The number of terms that arise is immense! We shall thus simply summarize the end result of very close case by case examination of all terms that arise (Chapters 13–15 of [**9**]).

Let us introduce the notation:

$$^{(X)}\tilde{\pi}|_{S_{\underline{u},u}} = {}^{(X)}i$$

and let $^{(X)}\hat{i}$ denote the trace free part.

We stress again that, in view of Section 9.4, deformation tensors are estimated from connection coefficients and obtain their appropriate δ-behavior from there.

9.10.1. *Multiplier borderline terms.* — For τ_m, borderline terms $e = 0$ occur only for $n = 1$ and $n = 3$.

For $n = 1$, we have the borderline terms

$$({}^{(K)}i, \alpha(W))\rho(W), \qquad ({}^{K}i \wedge \alpha(W))\sigma(W)$$

where we now use the α notation, etc., to describe the decomposition of an arbitrary Weyl field.

For $n = 3$, we have

$$({}^{(K)}i, \underline{\alpha}(W))\rho(W), \qquad ({}^{K}i \wedge \underline{\alpha}(W))\sigma(W).$$

9.10.2. *Commutation borderline terms.* — For $\overset{(n)}{\tau_c}$, borderline terms occur only in the case $n = 1$ and 3, and moreover, they only occur in ${}^{(X)}\overset{(1)}{\mathcal{J}}$.

In searching for borderline terms, one must be careful. The border line terms are not the principle terms. They are lower order (from the point of view of differentiability).

Let us define

$$\not{\mathcal{J}}(X, Y) = \mathcal{J}(X, \hat{L}, Y)$$

similarly $\underline{\not{\mathcal{J}}}$.

Let us decompose this as:

$$\not{\mathcal{J}} = \Theta - \Lambda \not{g} + K \not{\epsilon}.$$

In the case $n = 1$, we have $\Lambda\rho(W)$, $K\sigma(W)$.

In Λ, what gives rise to borderline terms is

$$\mathrm{tr}\underline{\chi}({}^{(X)}\hat{i}, \alpha),$$

whereas in K, it is the expression

$$\mathrm{tr}\underline{\chi}^{(X)}\hat{i} \wedge \alpha.$$

In the case $n = 3$, we have $(\underline{\Theta}, \underline{\alpha}(W))$, and the borderline terms in $\underline{\Theta}$ are

$$\mathrm{tr}\underline{\chi}(\rho^{(X)}\hat{i} - \sigma^{*(X)}\hat{i}).$$

9.11. The reductive structure

And here we have arrived at what is the most amazing and unexpected aspect of the structure!

The borderline terms as identified in Section 9.10 are cooperative with the stratification introduced in Section 9.7 in a remarkable way: Despite the fact that the system of inequalities satisfied by the $\overset{(n)}{\mathcal{E}}$ is still non-linear, it can be solved reductively so as to grow (modulo terms lower order in δ) sublinearly.

We have already seen that for $\overset{(0)}{\mathcal{E}}$, $\overset{(2)}{\mathcal{E}}$ there are no borderline terms, hence:

$$\overset{(0)}{\mathcal{E}} \leq \overset{(0)}{D} + 1, \tag{63}$$

$$\overset{(2)}{\mathcal{E}} \leq \overset{(2)}{D} + 1. \tag{64}$$

Since borderline terms for $\overset{(1)}{\mathcal{E}}$ (as identified in the previous section) are linear in α, this leads to an inequality of the form below:

$$\overset{(1)}{\mathcal{E}} \leq \overset{(1)}{D} + C\sqrt{\overset{(0)}{\mathcal{E}} + \overset{(1)}{\mathcal{E}}}\sqrt{\overset{(0)}{\mathcal{E}}}\sqrt{\overset{(2)}{\mathcal{E}}} + 1 \tag{65}$$

which while non-linear, is sublinear in $\overset{(1)}{\mathcal{E}}$, *given* $\overset{(0)}{\mathcal{E}}$, $\overset{(2)}{\mathcal{E}}$.

(As an exercise the reader may want to examine how terms α appear in the $\overset{(n)}{\mathcal{E}}$, in view of the definitions of Section 9.7).

Similarly,

$$\overset{(3)}{\mathcal{G}} \leq \overset{(3)}{D} + C\sqrt{\overset{(0)}{\mathcal{E}} + \overset{(1)}{\mathcal{E}}}\sqrt{\overset{(2)}{\mathcal{E}}}\sqrt{\overset{(3)}{\mathcal{G}}}. \tag{66}$$

The system closes!

We can also estimate now

$$\overset{(3)}{\mathcal{E}} \leq \overset{(3)}{D} + \ldots, \tag{67}$$

which has in fact decoupled from the rest.

We have bound the quantity $\mathcal{P}$ of (61)!

10. THE LOGIC OF THE PROOF

With the discussion of the bounds on the connection and on curvature in the above two sections, we have given an overview of all the main ideas from the point of view of analysis behind the proof of Theorem 3.1.

As discussed, however, already in Section 7, in the context of the *logic* of the proof of Theorem 3.1, the estimates of Sections 8 and 9 collectively represent only one (but by far the most important!) step of a *continuity argument*, where a region defined by a collection of bootstrap assumptions is successively enlarged. Specifically, the estimates of the above mentioned Sections 8 and 9 correspond to the step: "improving the bootstrap assumptions".

We turn in this section to discuss in more detail the structure of the actual continuity argument.

For the experts, we note that the continuity argument requires particular care because the framework for "local existence" is different from that used to prove estimates. This is handled in **[9]** in complete detail and with considerable technical artistry; in fact, the set-up used serves as a model for the careful treatment of these issues. For the reader's convenience, we have specific page references to the various steps of the argument.

10.1. The set $\mathscr{C}$

Let us recall the notation M_c of Section 3.1, for a general $c \in (u_0, -1]$.

Let us also recall the notation $\mathscr{P}$ from (61).

We begin with a definition

DEFINITION 10.1. — *Let data be fixed and let $\mathscr{C}$ be the set of real numbers $c \in (u_0, -1]$, such that*

1. *We have smooth solution in canonical coordinates on $M_c \setminus \Gamma_0$ and such that $M_0 \setminus \Gamma_0$ is Minkowski. Moreover for $c > u_0 + \delta$, the solution extends smoothly to the outer boundary $\underline{u} = \delta$.*
2. *A collection of boostrap assumptions[13] hold on M_c (necessary for Sobolev inequality, coercivity inequalities, and some additional ones necessary for comparison of energies).*
3.

$$\mathscr{P} \leq G(\overset{(0)}{D}, \overset{(1)}{D}, \overset{(2)}{D}, \overset{(3)}{D}) \tag{68}$$

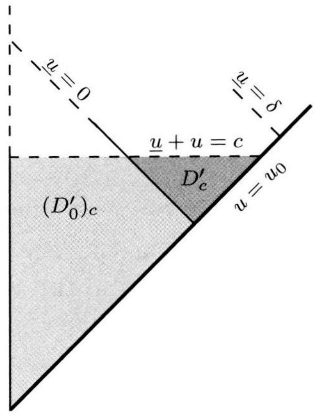
for a function G to be discovered[14] in the course of the proof.

---

[13] These are the bootstrap assumptions, proper, and are collected in Chapter 12, p 413 of **[9]**. We shall refer more generally to these assumptions *together with* (68) as "bootstrap assumptions".

[14] See formulas (16.16), (16.18) and (16.25) of **[9]**.

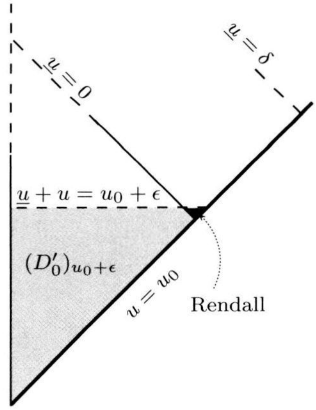
In Section 3.1, note that we depicted M_c only in the case $c \geq u_0 + \delta$. If $c < u_0 + \delta$, M_c is the region obtained by rotating the region D_c depicted below:



10.2. $\mathscr{A} \neq \varnothing$

Let us note that by appealing to Rendall's local well-posedness result, Theorem 1.2, one can solve the characteristic initial value problem for smooth initial data in a small neighborhood of S_{0,u_0}.

Using the implicit function theorem and compactness, then one can write the solution in the gauge of Section 3.1 in a region M_c for some $c > u_0$:



Moreover, the bootstrap assumptions corresponding to 2. in the definition of $\mathscr{A}$ as well as (68) can be seen to hold by continuity, for some $c > u_0$ (depending on the data). We have shown thus

PROPOSITION 10.2. — *$\mathscr{A}$ is non-empty.*

10.3. Semi-global existence

Set $c^* = \sup \mathscr{C}$. Theorem 3.1 in fact follows from the following:

THEOREM 10.3. — *If δ is suitably small depending on $\overset{(n)}{D}$, $n = 0,1,2,3,;$ $\mathcal{D}_0^\infty$, $\mathcal{D}_1^4$, $\mathcal{D}_2^4(\mathrm{tr}\underline{\chi})$, $\mathcal{D}_3(\mathrm{tr}\underline{\chi})$,*[15] *then $c^* = -1 \in \mathscr{C}$.*

The proof is by continuity. One shows that if $c \in \mathscr{C}$ with $c \leq 1$, then there exists an ϵ such that $c + \epsilon \in \mathscr{C}$.

Let us first describe the argument if $c > u_0 + \delta$, which is somewhat simpler.

1. The first (and by far most important and difficult!) step is to improve the bootstrap assumptions corresponding to 2. and the assumption (68). All together, this concerns Chapters 3–Chapter 16.1 of the work!

   More precisely, the logical order of this big step is as follows:

   (a) One first notes that, essentially since the bootstrap assumptions are defined by $\leq$, $c^* \in \mathscr{C}$ and the solution is defined in M_{c^*}. Given the bootstrap assumptions corresponding to 2., one can control the the coercivity and Sobolev constants allowing the energy fluxes defining $\mathscr{P}$ to be comparable to geometric L^2 bounds, etc. This is Chapters 10 and 11. (For instance, at this stage we can estimate $\mathscr{R}_0^\infty$ appearing on the right hand side of (43) in terms of $\mathscr{P}$.)

   (b) The estimates on the connection (Chapters 3–7, discussed here in Section 8) then apply in M_{c*}. The bounds on the connection then allow for control of the deformation tensor of the multiplier and commutation vector fields (Chapters 8–9). Moreover, as suggested for instance in the discussion of Section 8.3, the retrieved bounds are such that all the bootstrap assumptions of 2. can then be improved.

   (c) We now apply the divergence identity of Section 9.8 as in Section 9.8, and the reductive structure of the system of inequalities (63)–(67) allows us to improve inequality (68) for $\mathscr{P}$, say by a factor of 1/2, for proper choice of the function G. This is Chapters 12–16.1.

2. Because one will apply existence theorems that are most cleanly stated in the smooth category, it is necessary to first show uniform estimates in M_{c^*} for all higher norms. These estimates are now essentially linear. This is the entirety of Chapter 16.2.

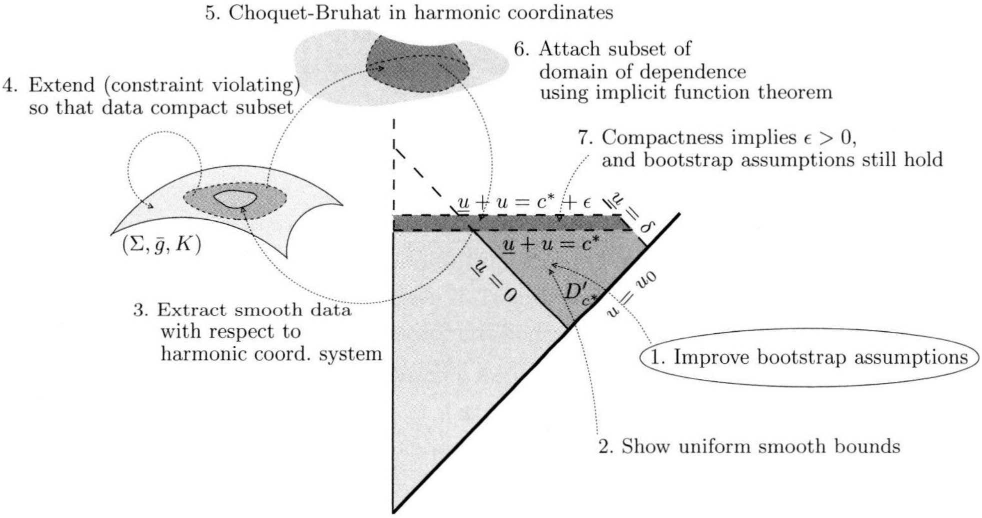
3. One now can use this to show that the solution extends smoothly to $\underline{u} + u = c^*$ and a smooth initial data set can be constructed for the reduced Einstein system. See Chapter 16.3, pp. 551–55.

---

[15] These are all initial data quantities and bounded by a positive continuous increasing function of M_8.

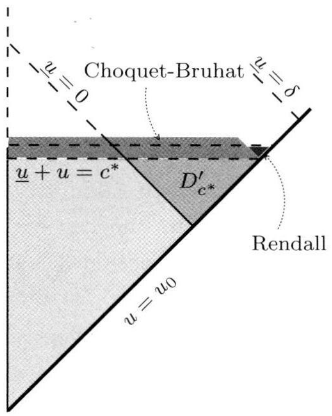
4. It is convenient to extend the data so that the interesting part of the data is contained in a compact set. The extension need not satisfy the constraints. See Chapter 16.3, pp. 555-56.
5. One then solves the reduced Einstein equations following Choquet-Bruhat (cf. Section 1.1.4). In the domain of dependence of the interesting part of the data, this is a genuine smooth solution of the Einstein equations as in Theorem 1.1. See Chapter 16.3, pp. 556–59.
6. Using the implicit function theorem, one can express a suitable part of this solution in the double null gauge of Section 3.1 and attach smoothly to M_{c^*}. See Chapter 16.3, pp. 559–69.
7. Compactness shows that this attachment covers some $M_{c^*+\epsilon}$, with $\epsilon > 0$, and since the bootstrap assumptions were improved in step 1, by continuity they still hold, for ϵ suitably small. See Chapter 16.3, pp. 570–71.

The above steps are depicted below:



Let us note that the earlier stages of the openness argument (corresponding to $c^* < u_0 + \delta$) require an extra step, as the curve $\underline{U} + u = c^*$ has not met the boundary $\underline{u} = \delta$.

This extra step is provided by appealing to Rendall's local existence Theorem 1.2 after step 6 in the above outline, just as in the proof of Proposition 10.2 above, and again appealing to the implicit function theorem and continuity for the construction of the gauge. This is illustrated below:



See Chapter 16.3, pp. 569–70.

11. THE KLAINERMAN-RODNIANSKI RELAXED HIERARCHY

Christodoulou's short pulse method has been revisited in a more recent and very insightful approach to the problem by Klainerman-Rodnianski [**27, 26**].

Let us make some very brief comments about this approach. The essential idea is to slightly relax the "short pulse hierarchy" of Section 5.5, to a less precise hierarchy, in a way which, though at first appears quite unnatural from the perspective of the behavior of ψ under the ansatz of Section 5.1, is suggested by a certain scaling property of the Einstein equations that the authors introduce.

At the level, say, of energy estimates for curvature as in Section 9, these scaling properties then provide a systematic approach to predicting the excess index e (see Section 9.9) of the non-linear terms which appear in the error terms τ (see (62)): For each arising term acquires a well-defined notion of *signature* which must be respected by the Einstein equations. This is a very important point, because, in practice, enumerating and analyzing all terms leading to the final conclusions summarized in Section 9.10 constitutes much of the work.

A family of relaxations of the original hierarchy are in fact possible. In this context, the authors first consider a certain *subcritical* hierarchy (with respect to their scaling). This hierarchy can be shown to propagate following the scheme outlined in Sections 8 and 9 above, *without having to enumerate carefully the structure of every nonlinear term arising in* τ, as scaling and signature arguments ensure $e > 0$ and thus the *absence* of borderline terms.

The propagation of this hierarchy allows one to state a semi-global existence theorem in the style of Theorem 3.1, for fixed u_0, again allowing large amplitude initial data. This can be used, for instance, to prove a theorem on the formation of *prescarred surfaces*, that is to say surfaces S such that $\mathrm{tr}\chi < 0$ only in some proper open subset $\mathcal{U} \subset S$.[16]

This subcritical scaling, however, is not compatible with the assumptions necessary for bona-fide trapped surface formation, which requires a large initial amplitude spread out over *all* directions of the sphere.

This difficulty can be overcome in the Klainerman-Rodnianski approach by slightly modifying the above relaxed hierarchy yet again, now allowing for certain violations of scaling. These are termed *anomalies.* As in Section 9.10, these anomalies will indeed give rise to borderline terms in the energy estimates, and one will need again to identify a reductive structure as in Section 9.11. There are far fewer such terms, however, and these are easier to systematically analyze.

An added simplification of the Klainerman-Rodnianski approach is that the whole argument is closed at one lower order of differentiability. In particular, one can cross off the entire lower line of the table of Section 9.6. This also leads to considerable simplifications, which of course translate to only requiring rougher control of the connection coefficients in Section 8.

There is of course, a price to be paid for this: Since the hierarchy is weaker and less geometrically motivated, one obtains less control of solutions than is obtained in the original approach. Nonetheless, control is sufficient to obtain the most basic geometrical conclusions.

These ideas have since been extended to the problem of trapped surface formation for the Einstein-Maxwell equations by P. Yu **[43]**. Yet another approach to Christodoulou's short pulse method applied to the vacuum equations, using the well-known first order symmetric hyperbolic reformulation due to H. Friedrich **[18]**, is given in **[37]**.

12. APPLICATIONS TO THE INCOMPLETENESS THEOREMS

The motivation for the notion of a trapped surface is its intimate relation with the incompleteness theorem, Theorem 0.1. In particular, we would certainly like to apply the result of Theorem 2.1 to derive analogous incompleteness statements for our own initial value problem. So far, we have only considered initial data up to advanced time $\underline{u} = \delta$. As the initial data is incomplete, the development will also be, but trivially so.

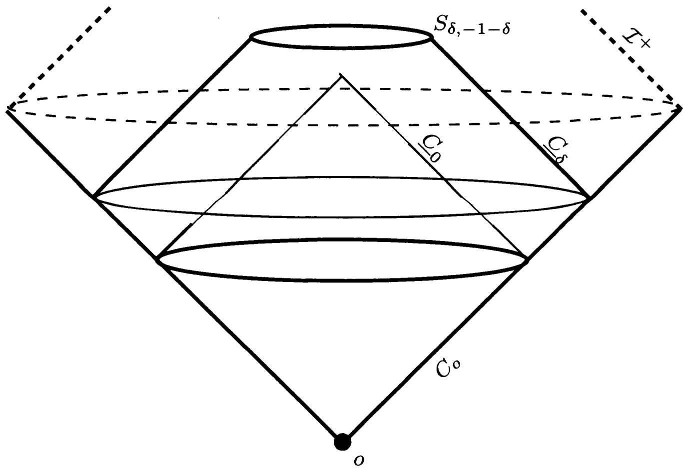
[16] In connection to such surfaces, see also the comments in the last paragraph of p. 581 of Chapter 17 of **[9]**.

To make a sensible statement about incompleteness, we must first complete our choice of initial data!

Fortunately, there is in fact an explicit construction of complete initial data (see Chapter 17, p. 579 of [**9**]) on C_o by extending ψ appropriately (note one must ensure that the data is regular—cf. Section 4.2.2). For such data, we note then an immediate corollary of Theorem 2.1 and Penrose's incompleteness theorem[(17)]:

Corollary 12.1. — *Let $(\mathcal{M}, g)$, be the maximal development of complete initial data on C_o, satisfying the assumption of Theorem 2.1. Then $(\mathcal{M}, g)$ is future causally geodesically incomplete.*

Given a suitable choice of complete data which is moreover *asymptotically flat* in a suitable sense, one can show that the solution exists for at least a finite retarded time (See [**10, 24**]). This allows one to define a notion of future null infinity $\mathcal{I}^+$ as in [**12, 10, 24**]:



The proof of Theorem 0.3 still applies in this context, and thus we obtain

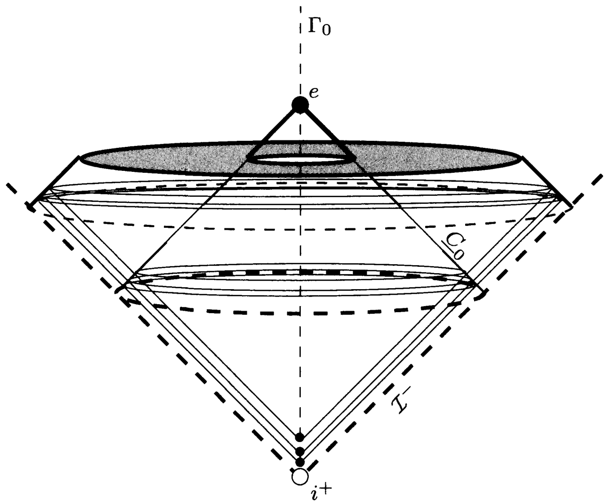
Corollary 12.2. — *Let $(\mathcal{M}, g)$, be the maximal development of complete initial data on C_o, satisfying the assumption of Theorem 2.1. Then $S_{\delta,-1-\delta} \cap J^-(\mathcal{I}^+) = \varnothing$. In particular, the spacetime $\mathcal{M}$ contains a black hole region $\mathcal{B}$ in the sense of definition (2).*

(17) This theorem also holds when the assumption of a Cauchy hypersurface is replaced by the assumption of having a complete null.

13. DATA AT PAST NULL INFINITY $\mathscr{I}^-$

The estimates of the hierarchy are such that one can immediately take the limit of the statement of Theorems 2.1 and 3.1 as $u_0 \to -\infty$.

Specifically, for a sequence of $u_0^i \to -\infty$, one considers a sequence of initial data on C_{u_0} defined by a fixed δ and fixed seed function ψ_0.



Applying the estimates of the hierarchy, noting that we have also estimates at all orders, it follows immediately by Arzela-Ascoli that there exists a subsequence which converges to a smooth solution of the Einstein equations.

Moreover, in view of the uniformity of the estimates in u_0, this limiting solution is such that one can attach a past boundary $\mathscr{I}^-$, to be thought of as past null infinity, and for which the limiting data $|u|^2 e(\underline{u}, u, \vartheta)$ converge to what is denoted in **[9]** by $e_\infty(\underline{u}, \vartheta)$ and which is just the past null infinity analogue of the square of the radiative amplitude $\Xi(u, \vartheta)$ defined in **[12]**. This has the interpretation of incoming radiative power per unit solid angle and its integral along the null generators of $\mathscr{I}^-$ for fixed ϑ thus corresponds to incoming radiative energy per unit solid angle (cf. the quantity (15) in the context of the discussion of the memory effect).

We thus obtain:

THEOREM 13.1. — *Let $0 < l < 1 < k$ be constants. Let us be given smooth asymptotic initial data at past null infinity which is trivial for advanced times $\underline{u} \leq 0$. Suppose that the incoming energy per unit solid angle in each direction in the advanced time interval*

$[0,\delta]$ is not less than $k/8\pi$. Then if δ is suitably small, the maximal development[18] *of the data contains a closed trapped surface S which is diffeomorphic to S^2 and has area* $\mathrm{Area}(S) \geq 4\pi l^2$.

See Chapter 17, pp. 580–81 of [**9**].

14. COMPLETENESS OF FUTURE NULL INFINITY AND APPROACH TO KERR

The given definition (2) of the notion of black hole, though intimately connected with Theorem 0.3, is in fact in reality too general.

For instance, if we imagine Minkowski space minus the future of two distinct points, say both on the Euclidean subset $\mathbb{R}^3 \subset \mathbb{R}^{3+1}$:

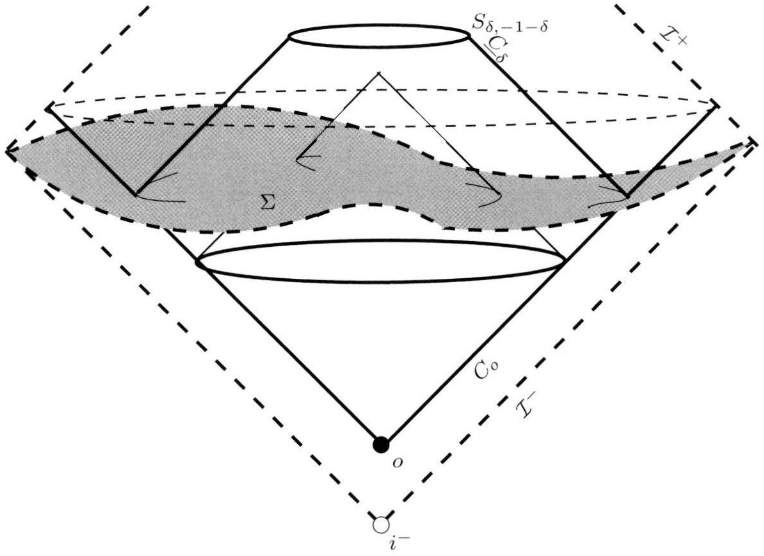
$$\mathbb{R}^{3+1} \setminus (J^+(p) \cup J^+(q)),$$

then this is a globally hyperbolic spacetime containing a black hole, according to definition (2). Of course, the above spacetime is not a maximal Cauchy development. On the other hand, one could easily imagine a maximal Cauchy development with two "first singularities", resulting in a causal geometry broadly similar to the above.

An additional distinguishing property of the future null infinity $\mathscr{I}^+$ attached to Schwarzschild or Kerr, is that $\mathscr{I}^+$ is *complete*. This essentially means that the null generators of $\mathscr{I}^+$ can be continued to both future and past to arbitrary values of affine parameter, or, alternatively, in the setup of Theorem 1.3, that it is naturally parametrized as $(-\infty,\infty) \times S^2$.

Now, if one takes the data to the limit on *past null infinity* $\mathscr{I}^-$ as in Section 13 above, then, using ideas from the stability of Minkowski space, one can easily show the *past* completeness of future null infinity $\mathscr{I}^+$ (and the existence of an asymptotically flat Cauchy surface Σ).

(18) We note that the uniqueness statement implicit in the characterization of the above as the maximal development of data prescribed at $\mathscr{I}^+$ follows from work of Friedrich [**18**].



Thus, in what follows, we need only consider the issue of *future* completeness of $\mathscr{I}^+$. Physically, this is the statement that asymptotic observers in the radiation zone can observe ad infinitum.

Now it is widely believed that if $\mathscr{I}^+$ is complete and $J^-(\mathscr{I}^+) \neq \varnothing$, then the metric in $J^-(\mathscr{I}^+)$ will eventually asymptote to the exterior of a collection of Kerr black holes, each rapidly moving away from the other. Here, however, we confront a fundamental difficulty. For, even if we were to *start* with Cauchy data *arbitrarily close* to Schwarzschild or Kerr, it is unknown whether the resulting $\mathscr{I}^+$ is complete.

Nonetheless we have the following conjecture

CONJECTURE 14.1 (Nonlinear stability of the Kerr family). — *Let* $|a| < M$,[19] *let* $(\mathcal{M}, g_{a,M})$ *denote the globally hyperbolic region of a subextremal Kerr manifold with two ends, and let* $\Sigma_{a,M}$ *denote a Cauchy hypersurface. Let* $(\Sigma, \bar{g}, K)$ *be vacuum initial data suitably close to Kerr data on* $\Sigma_{a,M}$. *Let* $(\mathcal{M}, \bar{g}, K)$ *denote its Cauchy development. Then one can attach to* $\mathcal{M}$ *an appropriate asymptotic boundary* $\mathscr{I}^+$ *(with two connected components!)* $\mathscr{I}^+_A$, $\mathscr{I}^+_B$, *such that both* $\mathscr{I}^+_A$, $\mathscr{I}^+_B$ *are complete and such that* g *asymptotes in* $J^-(\mathscr{I}^+_A)$ *and* $J^-(\mathscr{I}^+_B)$ *to nearby subextremal Kerr metrics* g_{a_A,M_A} *and* g_{a_B,M_B} *with* a_A, a_B *close to* a, *and* M_A, M_B *close to* M.

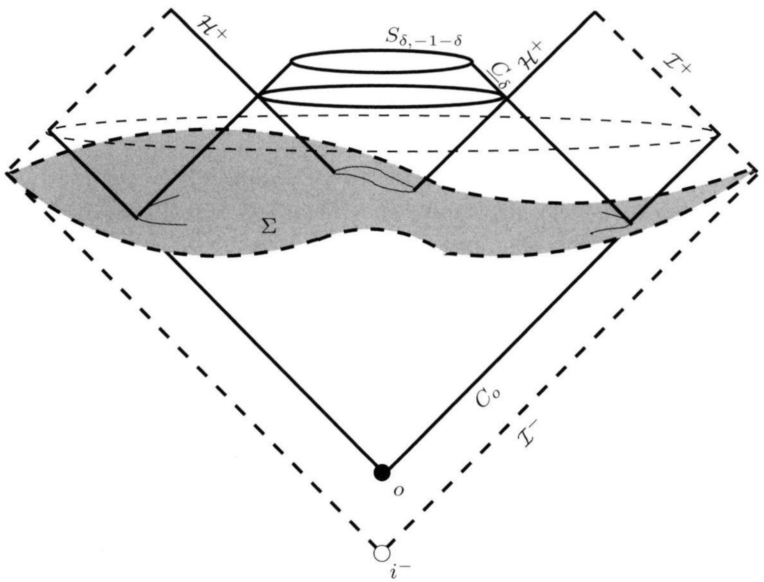
(19) The Kerr family depends on two parameters, *mass* M and *specific angular momentum* a, and the black hole case is $|a| \leq M$. The so called extremal case $|a| = M$ is special and is best excluded from the conjecture. See **[14, 13]** for background.

The above is more than just the statement of completeness of $\mathscr{I}^+$; it is the statement of asymptotic stability of the Kerr family. The point is, however, that, for supercritical quasilinear wave equations like the Einstein equations, these problems appear to be necessarily coupled. Thus, one does not expect to be able to prove the completeness property without understanding, quantitatively, the dispersion mechanism to a nearby Kerr.

Even the linear scalar aspects of the dispersive properties of waves on Kerr black hole backgrounds—a necessary prerequisite if Conjecture 14.1 is ever to be addressed, have only recently been understood. See [**14, 13**] for a survey and many references.

Returning now to the problem at hand, even though understanding the global properties of $J^-(\mathscr{I}^+)$ in the context of the initial data considered here is not strictly speaking a stability problem, it may be that a resolution of the above conjecture nonetheless opens the way to its solution as well, in conjunction of course with an extension of the short pulse hierarchy to the whole exterior region. One can thus also conjecture:

CONJECTURE 14.2. — *For suitable completions of the initial data on C_o (or $\mathscr{I}^-$) as in Section 12 (or 13), then future null infinity $\mathscr{I}^+$ is complete and the geometry in $J^-(\mathscr{I}^+)$ asymptotes to a subextremal Kerr geometry $g_{M,a}$.*



15. EPILOGUE: WEAK COSMIC CENSORSHIP

An even more ambitious problem than showing the completeness of null infinity for the particular case of data as in Conjectures 14.1 or 14.2 is the following

CONJECTURE 15.1 (Weak cosmic censorship). — *For generic asymptotically flat vacuum initial data, the maximal development possesses a complete future null infinity* $\mathscr{I}^+$.

This is one of the great open problems of classical relativity! In view of Theorem 0.1, one should view the above conjecture as the remaining analogue of the problem of *global existence* in general relativity.

In the original formulation of the above, due to Penrose, the caveat generic was not included. Christodoulou then showed **[5]** that for the analogue of the above problem for the self-gravitating scalar field under spherical symmetry, i.e. the study of spherically symmetric solutions to:

$$\mathrm{Ric}_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi \mathrm{T}_{\mu\nu} \tag{69}$$

$$\Box_g \psi = 0 \tag{70}$$

$$\mathrm{T}_{\mu\nu} = \partial_\mu\psi\partial_\nu\psi - \frac{1}{2}g_{\mu\nu}\partial^\mu\psi\partial_\nu\psi, \tag{71}$$

then there exist regular asymptotically flat initial data for which $\mathscr{I}^+$ is incomplete! (These examples satisfy also $\mathscr{B} = \varnothing$.) We say that collapse has formed *a naked singularity*.

Let us note that naked singularities have long been known to form when (69) is coupled to certain other type of matter models. For instance, again Christodoulou **[3]** in 1984, showed that the Oppenheimer-Snyder model of gravitational collapse of a homogeneous spherically symmetric ball of dust **[33]** is in some sense unstable to the formation of so-called shell-focusing naked singularities of infinite density.[20] It is widely thought that this behavior is an artifice of the unrealistic equation of state.

The case of the scalar field, however, is different. Scalar field matter ψ satisfies a linear equation (70) when g is considered frozen, so does not form singularities "on its own accord", and radiates to $\mathscr{I}^+$ just as with gravitational waves. Thus, the occurrence of naked singularities for the scalar field model (69)–(71) strongly suggests that similar phenomenon could happen in principle in the vacuum (1).

[20] This is somewhat ironic because the notion of black hole was first really accepted on the basis of this model!

Indeed, the main motivation for considering (69)–(71) is precisely as a "poor man's" vacuum equations (1) which can be studied in spherical symmetry.[21]

Remarkably, Christodoulou was able to show **[6]**—again in the context of the scalar field model (69)–(71) under spherical symmetry—that the previously constructed naked singularities **[5]** are unstable to perturbation of initial data, in fact *all naked singularities are unstable*, and that the analogue of Conjecture 15.1, as now formulated with the caveat generic, thus holds. In fact, something stronger is shown, namely that the set of data leading to an incomplete null infinity $\mathcal{I}^+$ have codimension at least 1 (in a suitable sense) in the space of all spherically symmetric data.

The key to proving Conjecture 15.1 for the (69)–(71) under spherical symmetry was proving that for such generic data, *all* "first singularities" are *preceded* by trapped surface formation. This notion of "first singularity" can be formalized in the language of terminal indecomposable past sets. See **[21]**. This motivates conjecturing a similar property for the vacuum:

CONJECTURE 15.2 (Trapped surface conjecture **[7]**). — *For generic asymptotically flat initial data $(\Sigma, \bar{g}, k)$ the maximal development $(\mathcal{M}, g)$ has the following property. If $\mathcal{P}$ is a terminal indecomposable past set with $\mathcal{P} \cap \Sigma$ of compact closure, then any open domain $\Sigma \supset \mathcal{D} \supset \overline{\mathcal{P} \cap \Sigma}$ has the property that the domain of dependence of $\mathcal{D}$ in $\mathcal{M}$ contains a closed trapped surface.*

The above conjecture would in fact imply (see **[7]**) a suitable formulation of Conjecture 15.1 for the Einstein vacuum equations (1).

It is of course a long way from the very specific setup of Theorem 2.1 to a general understanding of how and when trapped surfaces form. Again, however, the spherically symmetric case suggests that statements quite similar to Theorem 2.1 may be unexpectedly useful for the general problem. One should of course be weary of naively extrapolating from symmetric cases, where the very nature of the analysis is fundamentally simpler. On the other hand, the idea that a statement like Theorem 2.1, a large-data result for the vacuum equations, could be proven in the absence of symmetry assumptions would have seemed completely unrealistic only 5 years ago. Thus, the framework of Theorem 2.1 together with its revolutionary proof, while special, at the very least allow us to hope that, indeed, the above conjectures may one day be part of the realm of mathematical analysis.

[21] Note that Birkhoff's theorem (cf. **[21]**) says that spherically symmetric vacuum solutions are necessarily Schwarzschild.

16. ACKNOWLEDGEMENTS

This exposition would not have been possible in its current form without my having attended the series of lectures given by Demetrios Christodoulou at MSRI in September 2009, at a one-week "Hot topics workshop" dedicated in its entirety to [**9**]. Several sections of this exposition (e.g. Section 4.1, Sections 9.2–9.11) owe their general structure to his own presentation.

The presentation in Section 11 is in turn inspired by the beautiful talk of Igor Rodnianski at the same event.

I am grateful to both, not only for these lectures, but also for the many wonderful conversations we have had over the years. The point of view expressed in this exposition owes much to those discussions.

The above lectures are in fact available online at `http://www.msri.org/web/msri/scientific/workshops/hot-topics-workshops/show/-/event/Wm515` and are recommended in the strongest of terms for all who wish to learn more about this work.

17. APPENDIX

All concepts will refer to the double null foliation defined in Section 3.1.

17.1. The algebra calculus of S-tensor fields

S-tensor fields are defined as follows: ξ is an S 1-form on M if $\xi(L) = \xi(\underline{L}) = 0$. (Thus ξ is in fact specified by a smooth choice of 1 form on $S_{\bar{u},u}$ for each $S_{\bar{u},u}$.) A vector field V is an S-vector field if it is tangent to $S_{\bar{u},u}$). The notion then generalizes to higher order tensors.

Let us also define the following operation: Let θ be a 2-covariant S-tensor field. Let e_A be a local frame field for $S_{\underline{u},u}$. Define $\theta^\sharp$ by

$$(\theta^\sharp)_A^{\ B} = \theta_{AC}(\not{g}^{-1})^{CB}.$$

This is a T^1_1 S-tensor field. Note that

$$g(X, \theta^\sharp Y) = \theta(X, Y)$$

for any $X, Y \in T_p S_{\underline{u},u}$.

Now given two such objects θ, ϕ, define the 2-covariant S-tensor field $\theta \times \phi$ by

$$\theta \times \phi = g(\theta^\sharp, \phi^\sharp), \qquad \text{i.e.} \qquad (\theta \times \phi)_{AB} = \theta_{AC}\phi_B^{\ C}.$$

For two S 1-forms ξ, ξ', we denote by (ξ,ξ') the inner product of the 1-forms with respect to $\not g$,

$$(\xi,\xi') = \not g^{-1}(\xi,\xi') = \not g(\xi^\sharp,\xi'^\sharp), \qquad |\xi| = (\xi,\xi)^{1/2}$$

while we denote by $\xi\hat{\otimes}\xi'$ the symmetric trace-free 2-covariant S tensor field defined by

$$\xi\hat{\otimes}\xi' = \xi\otimes\xi' + \xi'\otimes\xi - (\xi,\xi')\not g.$$

Similarly, for a symmetric 2-covariant tensor field θ and an S 1-form ξ, we define

$$\theta\hat{\otimes}\xi = \theta\otimes\xi + \theta\tilde{\otimes}\xi - \theta^\sharp\cdot\xi\not g.$$

This is symmetric and trace free in the last two indices. The $\tilde{}$ denotes transposition.

17.1.1. *D and $\underline{D}$.* — We define the operators D, $\underline{D}$ on S-tensor fields. If X is an S-vector field, $\mathscr{L}_L X = [L,X]$ is also an S-vector-field. We define $DX = \mathscr{L}_L X$. On the other hand, if ξ is an S 1-form, then $D\xi$ is the restriction of $\mathscr{L}_L\xi$ to $TS_{\bar{u},u}$. Similarly for $\underline{D}$.

For a T^q_p S-tensor field θ, we denote by $D\theta$ the expression defined by considering θ on each C_u as extended to TC_u by the condition that it vanishes if one of the entries is L, and restricting to $TS_{\underline{u},u}$ of the Lie derivative of this object with respect to L.

Similarly we define $\underline{D}$.

Finally, we set $\hat{D}\theta$ and $\hat{\underline{D}}\theta$ to equal the trace-free parts of $D\theta$, $\hat{\underline{D}}\theta$.

17.1.2. $\mathrm{d}\!\!/\mathrm{iv}$, *$\nabla\!\!\!\!/$, $\mathrm{c}\!\!/\mathrm{url}$ and all that.* — For S-vector fields, $\mathrm{d}\!\!/\mathrm{iv}$, $\mathrm{c}\!\!/\mathrm{url}$ are self-explanatory, as is $\not d$ operating on arbitrary S-forms.

For S vector fields X, Y, we define $\nabla\!\!\!\!/_X Y = \Pi\nabla_X Y$, where Π is the projection to the tangent space of the surfaces.

For a two-covariant S tensor field θ, we define $\mathrm{d}\!\!/\mathrm{iv}\theta$ to be the S 1-form:

$$\mathrm{d}\!\!/\mathrm{iv}\theta_B = \nabla\!\!\!\!/_A \theta^{\sharp\ A}_B.$$

17.2. The connection coefficients

Recall from Section 3.1 the basic quantities at the level of the metric: $\not g$, Ω.

Recall also the three null frames of Section 3.1.2.

Here we define the connection coefficients $\chi, \underline{\chi}, \eta, \underline{\eta}, \omega, \underline{\omega}$.

17.2.1. *The second fundamental forms χ, $\underline{\chi}$.* — We define χ to be the intrinsic second fundamental form of C_u and $\underline{\chi}$ the intrinsic second fundamental form of $\underline{C}_{\underline{u}}$.

Specifically, χ is a symmetric 2-tensor field on C_u defined by

$$\chi(X,Y) = g(\nabla_X \hat{L}, Y).$$

Now, any vector $X \in T_p C_u$ may be decomposed as $PX + c\hat{L}$ where $PX \in T_p S_{\underline{u},u}$. We see easily that $\chi(X,Y) = \chi(PX,PY)$. This makes χ an S-tensor field.

It is useful to define also χ', $\underline{\chi}'$ where $\hat{L}$, $\hat{\underline{L}}$ are replaced by L', $\underline{L}$. Note that these are related by $\chi' = \Omega^{-1}\chi$, $\underline{\chi}' = \Omega^{-1}\underline{\chi}$.

17.2.2. *The torsion ζ (and torsion forms η, $\underline{\eta}$).* — We define an S 1-form ζ by

$$\zeta(X) = \frac{1}{2} g(\nabla_X \hat{L}, \hat{\underline{L}}), \qquad X \in T_p S_{\underline{u},u}.$$

Reversing the role of $\hat{L}$ and $\hat{\underline{L}}$, we define $\underline{\zeta}$. We see immediately that $\underline{\zeta} = -\zeta$.

For a given C_u, it is more natural to look at the torsion of $S_{\underline{u},u}$ with respect to the geodesic parametrization. We can define

$$\eta(X) = \frac{1}{2} g(\nabla_X L', \underline{L}), \qquad \underline{\eta}(X) = \frac{1}{2} g(\nabla_X \underline{L}', L).$$

We have

$$\eta = \zeta + \not{d} \log \Omega, \qquad \underline{\eta} = -\zeta + \not{d} \log \Omega.$$

Thus, if one controls both η, $\underline{\eta}$, one controls both ζ and $\not{d} \log \Omega$.

The geometric meaning of the torsion: $[\underline{L}, L]$ is the S tangential vector field:

$$[\underline{L}, L] = 4\Omega^2 \zeta^\sharp.$$

This is precisely the obstruction to integrability of the distribution spanned by $\underline{L}$, L, i.e. of the orthogonal complement of $T_p S_{\underline{u},u}$.

17.2.3. *The quantities ω, $\underline{\omega}$.* — We define finally

$$\omega = D \log \Omega, \qquad \underline{\omega} = \underline{D} \log \Omega.$$

17.3. The curvature in a null frame

The non-zero components of the curvature tensor of a Ricci flat metric are the following

$$\alpha(X,Y) = R(X, \hat{L}, Y, \hat{L}) \tag{72}$$

$$\beta(X) = \frac{1}{2} R(X, \hat{L}, \hat{\underline{L}}, \hat{L}) \tag{73}$$

$$\rho = \frac{1}{4}R(\hat{\underline{L}}, \hat{L}, \hat{\underline{L}}, \hat{L}) \tag{74}$$

$$\sigma \not\epsilon(X,Y) = R(\hat{\underline{L}}, \hat{L}, X, Y) \tag{75}$$

$$\underline{\beta}(X) = \frac{1}{2}R(X, \hat{\underline{L}}, \hat{\underline{L}}, \hat{L}) \tag{76}$$

$$\underline{\alpha}(X,Y) = R(X, \hat{\underline{L}}, Y, \hat{\underline{L}}) \tag{77}$$

where α and $\underline{\alpha}$ are trace free.

17.4. The structure equations of a null foliation

17.4.1. *The first variation equations.* — We have

$$D\not{g} = 2\Omega\chi, \qquad \underline{D}\not{g} = 2\Omega\underline{\chi}.$$

17.4.2. *The second variation equations.* — We have

$$D\chi' = \Omega^2\chi' \times \chi' - \alpha \tag{78}$$

$$\underline{D}\underline{\chi}' = \Omega^2\underline{\chi}' \times \underline{\chi}' - \underline{\alpha}. \tag{79}$$

17.4.3. *The torsion equations.* — We have

$$D\eta = \Omega(\chi^\sharp \cdot \underline{\eta} - \beta) \tag{80}$$

$$\underline{D}\underline{\eta} = \Omega(\underline{\chi}^\sharp \cdot \eta + \underline{\beta}) \tag{81}$$

$$\mathrm{c\not{u}rl}\underline{\eta} = \frac{1}{2}\chi \wedge \underline{\chi} - \sigma \tag{82}$$

$$\mathrm{c\not{u}rl}\eta = -\frac{1}{2}\chi \wedge \underline{\chi} + \sigma. \tag{83}$$

17.4.4. *The Codazzi equations.* — We have

$$\mathrm{d\not{i}v}\chi' - \not{d}tr\chi' + \chi'^\sharp \cdot \eta - \mathrm{tr}\chi'\eta = -\Omega^{-1}\beta \tag{84}$$

$$\mathrm{d\not{i}v}\underline{\chi}' - \not{d}tr\underline{\chi}' + \underline{\chi}'^\sharp \cdot \underline{\eta} - \mathrm{tr}\underline{\chi}'\underline{\eta} = \Omega^{-1}\underline{\beta}. \tag{85}$$

Note that the equations involving χ and $\underline{\chi}$ are not mutually coupled, while those involving η and $\underline{\eta}$ are. On the other hand, the latter are linear in η and $\underline{\eta}$.

17.4.5. *Propagation equations for ω, $\underline{\omega}$.* — We have

$$D\underline{\omega} = \Omega^2(2(\eta, \underline{\eta}) - |\eta|^2 - \rho) \tag{86}$$

$$\underline{D}\omega = \Omega^2(2(\eta, \underline{\eta}) - |\underline{\eta}|^2 - \rho). \tag{87}$$

17.4.6. *The Gauss equation.* — Let K denote the Gauss curvature of $\not{g}$. Then

$$K + \frac{1}{2}\mathrm{tr}\chi\mathrm{tr}\underline{\chi} - \frac{1}{2}(\chi, \underline{\chi}) = -\rho.$$

17.4.7. *The remaining propagation equations.* — These are:

$$\underline{D}(\Omega\chi) = \Omega^2(\not\nabla\eta + \widetilde{\not\nabla\eta} + 2\eta\otimes\eta + \frac{1}{2}(\chi\times\underline{\chi} + \underline{\chi}\times\chi) + \rho\not{g}) \tag{88}$$

$$D(\Omega\underline{\chi}) = \Omega^2(\not\nabla\underline{\eta} + \widetilde{\not\nabla\underline{\eta}} + 2\underline{\eta}\otimes\underline{\eta} + \frac{1}{2}(\chi\times\underline{\chi} + \underline{\chi}\times\chi) + \rho\not{g}). \tag{89}$$

17.5. The Einstein equations from the structure equations

The Einstein equations imply $\mathrm{tr}\alpha = 0$, thus, from (78), we obtain

$$D\mathrm{tr}\chi' = -\Omega^2|\chi'|^2. \tag{90}$$

This thus represents $\mathrm{Ric}(\hat{L}, \hat{L}) = 0$.

We can eliminate β, $\underline{\beta}$ in the torsion equations by using the Codazzi equations. The torsion equations then become the Einstein equations $\mathrm{Ric}(\hat{L}, X) = 0$, $\mathrm{Ric}(\hat{\underline{L}}, X) = 0$.

We may eliminate ρ from (86) using the Gauss equation to obtain:

$$\mathrm{Ric}(\hat{\underline{L}}, \hat{L}) = 0.$$

For the remaining Einstein equation $\mathrm{Ric}(X, Y) = 0$, for $X, Y \in T_pS_{\underline{u},u}$, we eliminate ρ from (88), using the Gauss equation.

17.6. Bianchi identities

$$\begin{aligned}
&\hat{\underline{D}}\alpha - \frac{1}{2}\Omega\mathrm{tr}\underline{\chi}\alpha + 2\underline{\omega}\alpha + \Omega\left\{-\not\nabla\hat{\otimes}\beta - (4\eta+\zeta)\hat{\otimes}\beta + 3\hat{\chi}\rho + 3^*\hat{\chi}\sigma\right\} = 0,\\
&\hat{D}\underline{\alpha} - \frac{1}{2}\Omega\mathrm{tr}\chi\underline{\alpha} + 2\omega\underline{\alpha} + \Omega\left\{\not\nabla\hat{\otimes}\underline{\beta} + (4\underline{\eta}-\zeta)\hat{\otimes}\underline{\beta} + 3\hat{\underline{\chi}}\rho - 3^*\hat{\underline{\chi}}\sigma\right\} = 0,
\end{aligned} \tag{91}$$

$$D\beta + \frac{3}{2}\Omega\mathrm{tr}\chi\beta - \Omega\hat{\chi}^\sharp\cdot\beta - \omega\beta - \Omega\left\{\mathrm{d\not{i}v}\alpha - (\underline{\eta}^\sharp + 2\zeta^\sharp)\cdot\alpha\right\} = 0,$$

$$\underline{D}\underline{\beta} - \frac{3}{2}\Omega\mathrm{tr}\underline{\chi}\underline{\beta} - \Omega\hat{\underline{\chi}}^\sharp\cdot\underline{\beta} - \underline{\omega}\underline{\beta} + \Omega\left\{\mathrm{d\not{i}v}\underline{\alpha} - (\eta^\sharp - 2\zeta^\sharp)\cdot\underline{\alpha}\right\} = 0,$$

$$\begin{aligned}
&\underline{D}\beta + \frac{1}{2}\Omega\mathrm{tr}\underline{\chi}\beta - \Omega\hat{\underline{\chi}}\cdot\beta + \underline{\omega}\beta - \Omega\left\{\not{d}\rho + {}^*\not{d}\sigma + 3\eta\rho + 3^*\eta\sigma + 2\hat{\chi}^\sharp\cdot\underline{\beta}\right\} = 0,\\
&D\underline{\beta} + \frac{1}{2}\Omega\mathrm{tr}\chi\underline{\beta} - \Omega\hat{\chi}\cdot\underline{\beta} + \omega\underline{\beta} + \Omega\left\{\not{d}\rho - {}^*\not{d}\sigma + 3\underline{\eta}\rho - 3^*\underline{\eta}\sigma + 2\hat{\underline{\chi}}^\sharp\cdot\beta\right\} = 0,
\end{aligned} \tag{92}$$

$$D\rho + \frac{3}{2}\Omega\mathrm{tr}\chi\rho - \Omega\left\{\mathrm{d\not{i}v}\beta + (2\underline{\eta}+\zeta, \beta) - \frac{1}{2}(\hat{\underline{\chi}}, \alpha)\right\} = 0,$$

$$\underline{D}\rho + \frac{3}{2}\Omega\mathrm{tr}\underline{\chi}\rho + \Omega\left\{\mathrm{d\not{i}v}\underline{\beta} + (2\eta-\zeta, \underline{\beta}) + \frac{1}{2}(\hat{\chi}, \underline{\alpha})\right\} = 0,$$

$$D\sigma + \frac{3}{2}\Omega\mathrm{tr}\chi\sigma + \Omega\left\{\mathrm{c\not{u}rl}\beta + (2\underline{\eta}+\zeta, {}^*\beta) - \frac{1}{2}\hat{\underline{\chi}}\wedge\alpha\right\} = 0,$$

$$\underline{D}\sigma + \frac{3}{2}\Omega\mathrm{tr}\underline{\chi}\sigma + \Omega\left\{\mathrm{c\not{u}rl}\underline{\beta} + (2\eta-\zeta, {}^*\underline{\beta}) + \frac{1}{2}\hat{\chi}\wedge\underline{\alpha}\right\} = 0.$$

17.7. Weyl fields and Weyl currents

A *Weyl field* on a general 4-dimension Lorentzian manifold $(\mathcal{M}, g)$ is a tensor field on $\mathcal{M}$ with the same algebraic properties of the Weyl curvature tensor:

$$W_{\beta\alpha\gamma\delta} = W_{\alpha\beta\delta\gamma} = -W_{\alpha\beta\gamma\delta}, W_{\alpha[\beta\gamma\delta]} = 0, (g^{-1})^{\mu\nu} W_{\mu\alpha\nu\beta} = 0.$$

Note that for a Weyl field W, we have that a left dual $*W$ and a right dual $W*$ coincide and again define Weyl fields. We have

$$\nabla^{\alpha} W_{\alpha\beta\gamma\delta} = \mathcal{J}_{\beta\gamma\delta}. \tag{93}$$

$\mathcal{J}_{\beta\gamma\delta}$ inherits algebraic properties from W which define the notion of *Weyl current.*

The above can be thought of as an analogue of (one part of) the Maxwell equations. In fact, for a Weyl field, the above is equivalent to

$$\nabla_{[\alpha,} W_{\beta\gamma]\delta\epsilon} = \epsilon_{\mu\alpha\beta\gamma} \mathcal{J}^{*}{}_{\mu\delta\epsilon}.$$

Note that if W is a Weyl field, then in general $\mathcal{L}_X W$ will not be a Weyl field because it will not be trace free. This can be remedied by introducing a modified $\tilde{\mathcal{L}}_X W$. This moreover satisfies $*\tilde{\mathcal{L}}_X W = \tilde{\mathcal{L}}_X * W$.

Similarly, for Weyl currents $\mathcal{J}$, we define $\mathcal{L}_X \mathcal{J}$ satisfying ${}^*\tilde{\mathcal{L}}_X \mathcal{J} = \tilde{\mathcal{L}}_X \mathcal{J}^*$.

Note the conformal property: Suppose on $(\mathcal{M}, g)$, $(W, \mathcal{J})$ is a Weyl pair. Then defining the rescaled metric $g' = \Omega^{-2} g$, $W' = \Omega^{-1} W$, $\mathcal{J}' = \Omega \mathcal{J}$, then $(W', \mathcal{J}')$ is a Weyl pair on $(\mathcal{M}, g')$.

As a consequence of the conformal property, we have

$$\nabla^{\alpha} (\tilde{\mathcal{L}}_X W)_{\alpha\beta\gamma\delta} = {}^{(X)}\mathcal{J}_{\beta\gamma\delta}(W)$$

where

$${}^{(X)}\mathcal{J}_{\beta\gamma\delta}(W) = (\tilde{\mathcal{L}}_X \mathcal{J})_{\beta\gamma\delta} + \sum_{i=1}^{3} {}^{(X)}\overset{(i)}{\mathcal{J}}_{\beta\gamma\delta}$$

and where

$${}^{(X)}\overset{(1)}{\mathcal{J}}_{\beta\gamma\delta} = \frac{1}{2} {}^{(X)}\tilde{\pi}^{\mu\nu} \nabla_{\mu} W_{\nu\beta\gamma\delta}, \qquad {}^{(X)}\overset{(2)}{\mathcal{J}}_{\beta\gamma\delta} = \frac{1}{2} {}^{(X)}P_{\mu} \nabla_{\mu} W^{\mu}{}_{\beta\gamma\delta}$$

$${}^{(X)}\overset{(3)}{\mathcal{J}}_{\beta\gamma\delta} = \frac{1}{2} \left({}^{X}q_{\mu\beta\gamma} W^{\mu\nu}{}_{\gamma\delta} + {}^{X}q_{\mu\gamma\nu} W^{\mu}{}_{\beta}{}^{\nu}{}_{\delta} + {}^{X}q_{\mu\delta\nu} W^{\mu}{}_{\beta\gamma}{}^{\nu} \right).$$

Remarkably, q is again a Weyl current.

REFERENCES

[1] L. Bieri & N. Zipser – *Extensions of the stability theorem of the Minkowski space in general relativity*, AMS/IP Studies in Advanced Math., vol. 45, Amer. Math. Soc., 2009.

[2] Y. Choquet-Bruhat & R. Geroch – "Global aspects of the Cauchy problem in general relativity", *Comm. Math. Phys.* **14** (1969), p. 329–335.

[3] D. Christodoulou – "Violation of cosmic censorship in the gravitational collapse of a dust cloud", *Comm. Math. Phys.* **93** (1984), no. 2, p. 171–195.

[4] ______, "Nonlinear nature of gravitation and gravitational-wave experiments", *Phys. Rev. Lett.* **67** (1991), no. 12, p. 1486–1489.

[5] ______, "Examples of naked singularity formation in the gravitational collapse of a scalar field", *Ann. of Math.* **140** (1994), no. 3, p. 607–653.

[6] ______, "The instability of naked singularities in the gravitational collapse of a scalar field", *Ann. of Math.* **149** (1999), no. 1, p. 183–217.

[7] ______, "On the global initial value problem and the issue of singularities", *Classical Quantum Gravity* **16** (1999), no. 12A, p. A23–A35.

[8] ______, "The global initial value problem in general relativity", in *9th Marcel Grosmann Meeting (Rome 2000)*, World Sci. Publishing, 2002, p. 44–54.

[9] ______, *The formation of black holes in general relativity*, EMS Monographs in Math., European Mathematical Society (EMS), Zürich, 2009.

[10] ______, "Notes on the geometry of null hypersurfaces".

[11] ______, "Topics on black holes and gravitational waves", lectures held at ETH-Zürich, 2011.

[12] D. Christodoulou & S. Klainerman – *The global nonlinear stability of the Minkowski space*, Princeton Mathematical Series, vol. 41, Princeton Univ. Press, 1993.

[13] 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*, World Sci. Publ., Hackensack, NJ, 2010, p. 421–432.

[14] ______, "Lectures on black holes and linear waves", to be published by Clay Mathematics Institute.

[15] T. de Donder – "La gravifique einsteinienne", *Annales de l'Observatoire Royal de Belgique* (1921).

[16] A. Einstein – "Über Gravitationswellen", *Königlich Preussiche Akademie der Wissenschaften Berlin Sitzungberichte* (1918), p. 154–167.

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

[18] H. Friedrich – "The asymptotic characteristic initial value problem for Einstein's vacuum field equations as an initial value problem for a first-order quasilinear symmetric hy", *Proc. Roy. Soc. London Ser. A* **378** (1981), no. 1774, p. 401–421.

[19] K. Friedrichs & H. Lewy – "Über die Eindeutigkeit und das Abhängigkeitsgebiet der Lösungen beim Anfangswertproblem linearer hyperbolischer Differentialgleichungen", *Math. Ann.* **98** (1928), no. 1, p. 192–204.

[20] H. Müller zum Hagen – "Characteristic initial value problem for hyperbolic systems of second order differential equations", *Ann. Inst. H. Poincaré Phys. Théor.* **53** (1990), no. 2, p. 159–216.

[21] S. W. Hawking & G. F. R. Ellis – *The large scale structure of space-time*, Cambridge Univ. Press, 1973, Cambridge Monographs on Mathematical Physics, No. 1.

[22] D. Kennefick – "Controversies in the history of the radiation reaction problem in general relativity", in *The expanding worlds of general relativity (Berlin, 1995)*, Einstein Stud., vol. 7, Birkhäuser, 1999, p. 207–234.

[23] S. Klainerman – "Uniform decay estimates and the Lorentz invariance of the classical wave equation", *Comm. Pure Appl. Math.* **38** (1985), no. 3, p. 321–332.

[24] S. Klainerman & F. Nicolò – *The evolution problem in general relativity*, Progress in Mathematical Physics, vol. 25, Birkhäuser, 2003.

[25] S. Klainerman & I. Rodnianski – "Rough solutions of the Einstein-vacuum equations", *Ann. of Math.* **161** (2005), no. 3, p. 1143–1193.

[26] ______, "On emerging scarred surfaces for the Einstein vacuum equations", *Discrete Contin. Dyn. Syst.* **28** (2010), no. 3, p. 1007–1031.

[27] ______, "On the formation of trapped surfaces", *Acta Math.* **208** (2012), no. 2, p. 211–333.

[28] G. Lemaître – "L'univers en expansion", *Publication du Laboratoire d'Astronomie et de Géodésie de l'Université de Louvain* **9** (1932), p. 171–205.

[29] J. Leray – *Hyperbolic differential equations*, The Institute for Advanced Study, Princeton, N. J., 1953.

[30] H. Lindblad & I. Rodnianski – "The global stability of Minkowski space-time in harmonic gauge", *Ann. of Math.* **171** (2010), no. 3, p. 1401–1477.

[31] J. Luk – "On the local existence for the characteristic initial value problem in general relativity", *Int. Math. Res. Not.* **2012** (2012), no. 20, p. 4625–4678.

[32] E. Newman & R. Penrose – "An approach to gravitational radiation by a method of spin coefficients", *J. Mathematical Phys.* **3** (1962), p. 566–578.

[33] J. R. OPPENHEIMER & H. SNYDER – "On continued gravitational contraction", *Phys. Rev.* **56** (1939), p. 455–459.

[34] R. PENROSE – "Gravitational collapse and space-time singularities", *Phys. Rev. Lett.* **14** (1965), p. 57–59.

[35] F. A. E. PIRANI – "Invariant formulation of gravitational radiation theory", *Phys. Rev.* **105** (1957), p. 1089–1099.

[36] F. PLANCHON & I. RODNIANSKI – "On uniqueness for the Cauchy problem in general relativity", preprint.

[37] M. REITERER & E. TRUBOWITZ – "Strongly focused gravitational waves", *Comm. Math. Phys.* **307** (2011), no. 2, p. 275–313.

[38] A. D. RENDALL – "Reduction of the characteristic initial value problem to the Cauchy problem and its applications to the Einstein equations", *Proc. Roy. Soc. London Ser. A* **427** (1990), no. 1872, p. 221–239.

[39] J. SCHAUDER – "Das Anfangswertproblem einer quasilinearen hyperbolischen Differentialgleichung zweiter Ordnung in beliebiger Anzahl yon unabhängigen Veränderlichen", *Fundam. Math.* **24** (1935).

[40] K. STELLMACHER – "Zum Anfangswertproblem der Gravitationsgleichungen", *Math. Ann.* **115** (1938), no. 1, p. 136–152.

[41] K. S. THORNE – "Gravitational-wave bursts with memory: the Christodoulou effect", *Phys. Rev. D* **45** (1992), no. 2, p. 520–524.

[42] R. M. WALD – *General relativity*, University of Chicago Press, 1984.

[43] P. YU – "Dynamical formation of black holes due to the condensation of matter field", preprint arXiv:1105.5898.

Mihalis DAFERMOS

D.P.M.M.S.
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB – UK
E-mail : M.Dafermos@dpmms.cam.ac.uk